

# UNIQUENESS OF SURFACE DIAGRAMS OF SMOOTH 4-MANIFOLDS

JONATHAN D. WILLIAMS

ABSTRACT. In [W] there appeared a new way to specify any smooth closed 4-manifold by an orientable surface decorated with simple closed curves. These curves are cyclically indexed, and each curve has a unique transverse intersection with the next. These are called surface diagrams, and each comes from a certain type of map from the 4-manifold to the two-sphere. The aim of this paper is to give a uniqueness theorem stating that these objects (coming from maps within a fixed homotopy class) are unique up to four moves: stabilization, handleslide, multislide, and shift.

## CONTENTS

1. Introduction	1
1.1. Background	2
2. Maps and diagrams	3
2.1. Critical points	3
2.2. Surface diagrams	3
3. Moves on surface diagrams	5
3.1. Handleslide	5
3.2. Stabilization	8
3.3. Multislide	10
3.4. Shift	11
4. Proof of Theorem 1	12
4.1. Splicing cusp arcs	13
4.2. Modification of the critical manifold	14
4.3. Modification of the immersion of the critical manifold	22
5. Appendix: validity of isotopies	34
References	36

## 1. INTRODUCTION

This paper is concerned with a certain class of maps from smooth closed orientable 4-manifolds to the two-sphere, called *simplified purely wrinkled fibrations* in [W], and their so-called *base diagrams*. Such a map  $\alpha_0 : M \rightarrow S^2$  defines a *surface diagram*  $(\Sigma, \Gamma)$ , where  $\Sigma$  is a closed, orientable surface decorated with a collection  $\Gamma$  of simple closed curves. Similar to a Heegaard diagram of a 3-manifold, a surface diagram specifies  $M$  up to diffeomorphism. Section 3 describes certain moves, called *handleslide*, *stabilization*, *multislide* and *shift*, that can be performed within the class of surface diagrams for a fixed 4-manifold; these moves are all induced

by particular homotopies of the fibration map  $\alpha_0$ . Section 4 proves the following theorem:

**Theorem 1.** Suppose  $\alpha_0, \alpha_1: M \rightarrow S^2$  are two homotopic simplified purely wrinkled fibrations with surface diagrams  $(\Sigma^i, \Gamma^i)$ ,  $i = 0, 1$ . Then there is a finite sequence of surface diagrams, beginning with  $(\Sigma^0, \Gamma^0)$  and ending with  $(\Sigma^1, \Gamma^1)$ , obtained by performing stabilizations, handleslides, shifts, multislides, and their inverses. When  $M$  is simply connected, the sequence can be assumed to be free of multislides.

**1.1. Background.** The subject of simplified purely wrinkled fibrations arose from the study of broken Lefschetz fibrations on smooth 4-manifolds, which were first introduced in [ADK] to generalize the correspondence between Lefschetz fibrations and symplectic 4-manifolds up to blowup (see [GS] for further details on symplectic structures and Lefschetz fibrations). Roughly, the result of [ADK] was that if one allows  $\omega$  to vanish in a controlled way along an embedded one-submanifold of  $M$  (that is,  $(M, \omega)$  is a *near-symplectic* 4-manifold), then a suitable blowup of  $(M, \omega)$  has a *broken* Lefschetz fibration such that, away from its vanishing locus,  $\omega$  restricts to a volume form on the fiber. Discussed below, broken Lefschetz fibrations are a mild generalization of Lefschetz fibrations in which the critical locus is allowed to contain a one-submanifold of critical points; when the fibration corresponds to a near-symplectic form  $\omega_{ns}$ , this critical one-submanifold coincides with its vanishing locus of  $\omega_{ns}$ ; see for example [B1, L1]. Though mild, this generalization greatly increases the collection of 4-manifolds that admit such a fibration structure: it is known that *every* smooth, orientable 4-manifold admits a broken Lefschetz fibration (though it may not correspond to any near-symplectic form); see for example [AK, B2, L1].

The study of broken Lefschetz fibrations is partly motivated by the effort to understand the Seiberg-Witten invariants of a smooth 4-manifold  $M$  geometrically. When  $b^{2+}$  is positive, the Seiberg-Witten invariants at their most basic level define a map from  $H^2(M; \mathbb{Z})$  to the integers, defined as an algebraic count of solutions to a nonlinear elliptic pair of partial differential equations on  $M$  [M]. In a 1996 paper, Taubes showed that, for symplectic 4-manifolds, solutions to the Seiberg-Witten equations correspond to pseudoholomorphic curves which contribute to a special Gromov invariant he defined, called  $Gr$ . Pseudoholomorphic curves are submanifolds of two real dimensions that are singled out by the chosen symplectic structure and other auxiliary data (these are choices which are later shown to not affect the values of  $Gr$ ) [T1].

Recent efforts to geometrically recover Seiberg-Witten theory for 3-manifolds and more general 4-manifolds have revolved around equipping a given manifold with structures that resemble surface bundles, namely Morse functions in three dimensions and (broken) Lefschetz fibrations in four dimensions. An initial development in the latter case appeared in a 2003 paper in which Donaldson and Smith defined a *standard surface count* for symplectic 4-manifolds, counting pseudoholomorphic curves which are sections of a fiber bundle associated to a Lefschetz fibration [DS]. Soon after, Usher showed that the standard surface count is equivalent to  $Gr$ , and thus the Seiberg-Witten invariant [U]. Though this equivalence is only known to hold for a suitably chosen Lefschetz fibration whose existence is equivalent to the existence of a symplectic form, it gave a promising inroad to generalization: broken Lefschetz fibrations offer a way to continue their approach into

nonsymplectic territory, further guided by an existence result of [T2] stating that if a near-symplectic 4-manifold has nonvanishing Seiberg-Witten invariant, then there is a pseudoholomorphic subvariety with boundary given by the vanishing locus of  $\omega$ . In a 2007 paper, Perutz defined a generalization of the standard surface count for near-symplectic broken Lefschetz fibrations, called the *Lagrangian matching invariant*, that fits in nicely with experts' expectations and [T2] in numerous ways. For example, there are formal similarities, vanishing theorems and calculations in special cases that coincide with those for the Seiberg-Witten invariant. However, it remains to show that it is a smooth invariant of the underlying 4-manifold (not just the isomorphism class of its chosen fibration structure), and it is generally difficult to compute. A large amount of the motivation for studying surface diagrams is the possibility that they will prove useful in addressing this pair of issues. *Acknowledgments.* The author would like to thank Denis Auroux and Rob Kirby for exceedingly helpful input during the preparation of this work, which was supported by NSF grant DMS-0635607.

## 2. MAPS AND DIAGRAMS

**2.1. Critical points.** Let  $M$  be a smooth closed 4-manifold. A generic map  $M \rightarrow S^2$  resembles a surface bundle, except one allows the presence of a one-dimensional critical locus that is a union of *fold* and *cuspl* points. The fold locus is an embedded one-submanifold of  $M$  with the following local model at each point:

$$(1) \quad (x_1, x_2, x_3, x_4) \mapsto (x_1, x_2^2 + x_3^2 \pm x_4^2).$$

When the sign above is negative, it is known variously as an *indefinite fold*, *round singularity*, or *broken singularity* depending on context. Figure 1a is a picture of the target space, depicting the fibration structure nearby. In that figure, like the one to its right, what appears is the target disk of a map  $D^4 \rightarrow D^2$ , with bold arcs representing the image of the critical locus. A surface is pictured in the region of regular values that have that surface as their preimage, and tracing point preimages above a horizontal arc from left to right gives the foliation of  $\mathbb{R}^3$  by hyperboloids, first one-sheeted, then two-sheeted, with a double cone above the fold point. The circle (drawn on the cylinder to the left) that shrinks to the cone point is called the *round vanishing cycle* for that arc in the context of broken Lefschetz fibrations. When dealing with purely wrinkled fibrations there are no Lefschetz vanishing cycles, so the term *vanishing cycle* will suffice in that case.

Each cuspl point is a common endpoint of two open arcs of fold points, with the local model

$$(2) \quad (x_1, x_2, x_3, x_4) \mapsto (x_1, x_2^3 - 3x_1x_2 + x_3^2 \pm x_4^2).$$

When the sign above is negative, it is called an *indefinite cuspl* and is adjacent to two indefinite fold arcs as in Figure 1b: here two fold arcs meet at a cuspl point, for which the two vanishing cycles must intersect at a unique point in the fiber.

**2.2. Surface diagrams.** For the purposes of this paper, a broken Lefschetz fibration is simply a smooth map from a 4-manifold  $M$  to a surface  $F$  (usually the sphere or the disk) whose critical locus is a collection of Lefschetz critical points and indefinite folds, and a purely wrinkled fibration is a stable map  $M \rightarrow F$  whose critical locus is free of definite folds. In other words, the critical set of a purely wrinkled fibration is a union of indefinite cusps and indefinite folds. For details on

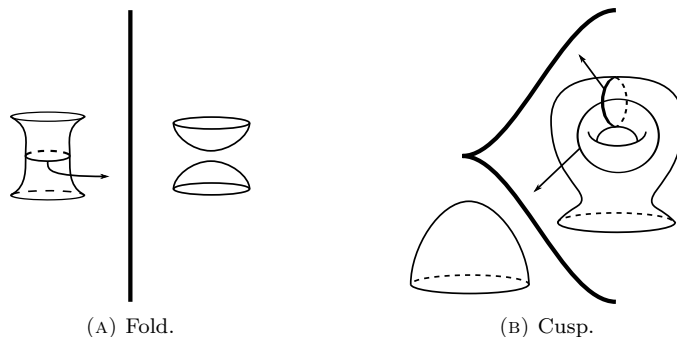


FIGURE 1. Critical points of purely wrinkled fibrations.

stability and the critical loci of stable maps, see [Wa, W]. By Corollary 1 of [W], every broken Lefschetz fibration can be modified by a (possibly long) sequence of moves (which are chosen from a short list initially appearing in [L1]) into a purely wrinkled fibration whose critical locus is one circle embedded by the fibration map into  $S^2$ ; combining terminology from [B1] and [L1], such a map is called a *simplified purely wrinkled fibration* (in fact, the proof of that corollary implies this can be done for *any* map  $M \rightarrow S^2$ ). Such a map  $f: M \rightarrow S^2$  has an orientable genus  $g$  surface as regular fiber over one of the disks that comprise  $S^2 \setminus f(\text{crit } f)$ , and the generic fiber over the other disk has genus  $g - 1$ ; choosing a regular value  $p$  in the higher-genus side and reference arcs from  $p$  to the various indefinite fold arcs, one obtains a collection of simple closed curves in  $\Sigma_g$  by recording the round vanishing cycles in a counter-clockwise direction going around the cusped critical circle. For this reason, the curves are relatively indexed by  $\mathbb{Z}/k\mathbb{Z}$ , where  $k$  is the number of fold arcs.

**Definition 1.** Assuming  $g \geq 3$ , a *surface diagram*  $(\Sigma_g, \Gamma)$  of  $M$  is the higher-genus fiber  $\Sigma_g$ , decorated with the relatively  $\mathbb{Z}/k\mathbb{Z}$ -indexed collection  $\Gamma = (\gamma_1, \dots, \gamma_k)$  of round vanishing cycles, up to orientation-preserving diffeomorphism.

**Remark 1.** The requirement that  $g \geq 3$  (which can always be satisfied by applying the modification discussed in Section 3.2 below) is necessary for the following reason: after using  $(\Sigma_g, \Gamma)$  to form the fibration consisting of the preimage of a neighborhood of the critical image, if  $g = 1$  or  $g = 2$  there are various ways to close off the higher- and lower-genus sides with a copies of  $(\text{fiber}) \times D^2$ , according to the elements of  $\pi_1(\text{diff}(S^2)) = \mathbb{Z}_2$  and  $\pi_1(\text{diff}(T^2)) = \mathbb{Z}^2$ , where  $\text{diff}(\Sigma)$  is the diffeomorphism group of  $\Sigma$  (see [ADK, B1, H] for explicit examples). Since  $\text{diff}(\Sigma_g)$  is simply connected for  $g \geq 2$ , a surface diagram as defined specifies the total space of the fibration up to orientation-preserving diffeomorphism.

**Remark 2.** The local model for the cusp requires the two round vanishing cycles to transversely intersect at a unique point in the fiber; for this reason, consecutive elements of  $\Gamma$  must intersect in this way. Also, each element of  $\Gamma$  must be an embedded circle in  $\Sigma_g$ . These two facts combine to form what we call the *intersection condition for surface diagrams*, or simply *the intersection condition* when context makes it clear.

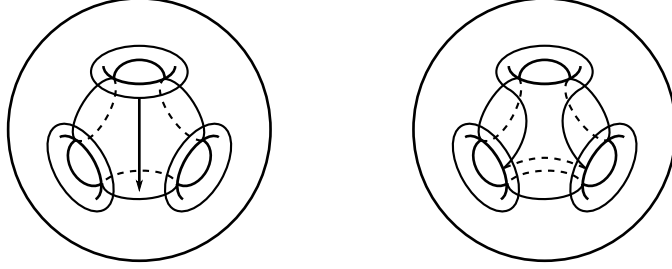


FIGURE 2. Sliding one round vanishing cycle over another.

### 3. MOVES ON SURFACE DIAGRAM

This section describes discrete moves that may be used to modify a surface diagram. These moves preserve the diffeomorphism type of the 4-manifold specified by the surface diagram because they come from the endpoints of deformations.

**Definition 2.** For purely wrinkled fibrations  $\alpha_0, \alpha_1$ , a *deformation*  $\alpha = \alpha_t$ ,  $t \in [0, 1]$  is a homotopy realized by a sequence of the moves in section 2.3 of [W].

To keep the distinction clear, the terms *handleslide*, *stabilization*, etc. will always refer to the move on surface diagrams, while the terms *handleslide deformation* and *stabilization deformation* will refer to the corresponding modifications of fibration maps.

**3.1. Handleslide.** In a Heegaard diagram for a 3-manifold, the attaching circles can be considered vanishing cycles for critical points of an associated Morse function; see e.g. [L2] for an application of this viewpoint. In a surface diagram, the handleslide appears exactly like that of Heegaard diagrams and comes, in some sense, from a one-parameter family of Heegaard handleslides. The difference for surface diagrams is that there is no chosen partition of the round vanishing cycles into sets of pairwise disjoint circles and there is no obvious linear independence requirement for the elements of  $\Gamma$  in  $H_1(\Sigma)$ , leading to additional assumptions that are necessary and sufficient to preserve the intersection condition.

Within the fibration underlying a surface diagram, denote four distinct components of the fold locus by  $A, B, C, D$  (with  $A, B, C$  consecutive and with vanishing cycles  $a, b, c, d$ , respectively). Suppose the remaining vanishing cycle  $d$  is disjoint from each of  $a, b$  and  $c$  (note this implies  $[a]$  and  $[d]$  are linearly independent in the homology of the surface because of the intersection condition for  $a, b, c$ ). The handleslide then consists of replacing  $b$  with any simple closed curve  $b'$  such that  $b, b', d$  bound an embedded thrice-punctured sphere  $S$  in the surface. The modification comes from the following deformation: in base diagrams,  $B$  (and a small amount of  $A$  and  $C$ ) moves across the higher-genus region past  $D$  in an  $R_2$  move. This is possible by the disjointness condition. In this lower-genus area the vanishing cycle for  $A$  becomes able to slide over either of the points that become identified as one traces a reference arc to  $D$ ; in other words, the boundary component of  $S$  coming from  $D$  is capped off in this lower-genus region. The last step is to cancel the intersections of  $A$  and  $D$  by the reverse  $R_2$  move in which  $A$  moves back across  $D$  into the higher-genus region. The vanishing cycles have now changed by a band sum of  $a$  and  $d$ , giving the embedding of  $S$ . Figure 2 gives an example. Note that

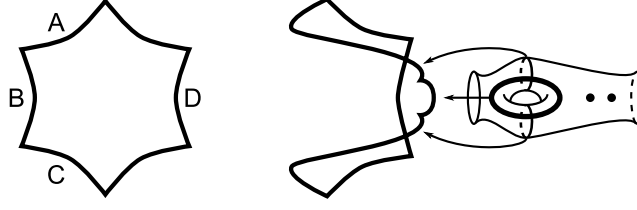


FIGURE 3. Base diagrams for part of the handleslide deformation, which is an isotopy move in which the vanishing cycle for  $B$  slides over that of  $D$ . In the fiber on the right, the bold vanishing cycle wanders over one of the two points before the two intersections are canceled by an  $R_2$  move.

the disjointness condition is sufficient for the resulting surface diagram to satisfy the intersection condition; in other words, for the second  $R_2$  move to be valid.

**Remark 3.** For each of the moves in Theorem 1, the highest-genus region of a base diagram momentarily becomes disconnected. Initially, reference fibers chosen over various points in the higher-genus region are decorated with identical sets of vanishing cycles up to isotopy. Depending on what goes on while there are two higher-genus components, there can be some variation in how the reference fibers over points in each component become identified when the components reunite, and this is the subject of the remark. For now, we restrict attention to the a deformation like in Figure 3 as follows.

**Lemma 1.** Suppose an SPWF experiences an  $R_2$  move in which two intersection points are formed by two fold arcs approaching each others' higher-genus sides, only to be cancelled by a reverse  $R_2$  move. Then the surface diagrams before and after such a deformation are related by some sequence of handleslides.

*Proof.* In this situation, it is useful to consider in the initial SPWF a reference fiber  $\Sigma_g$  over a point just to the lower-genus side of a fold arc  $\gamma$ , along with a short *reference path* across  $\gamma$  into its higher-genus side. As one traces fibers above this path, two points  $p, p'$  become identified in a surgery that increases the genus of the fiber by one as in Figure 1a. Then the endpoint of this path gives a reference fiber  $\Sigma_{g+1}$  to which one can add the vanishing cycles for folds bounding that higher-genus region. Pulling this picture back across  $\gamma$ , all the vanishing cycles descend to circles in  $\Sigma_g$  except those that intersect the vanishing cycle of  $\gamma$ ; these instead appear as *vanishing arcs* whose endpoints are  $p$  and  $p'$ . The ends of these arcs come equipped with a bijection  $\beta$  between those at  $p$  and those at  $p'$  by the way they pair up on either side of the vanishing cycle of  $\gamma$  in  $\Sigma_{g+1}$ : for this reason, the union of vanishing arcs obtained by picking a point on one vanishing arc, following it to one end  $e$ , continuing at  $\beta(e)$  and on until returning to the chosen point will be called a vanishing cycle just like the other simple closed curves. In this way, the base diagram can be recovered from a reference fiber in the lower-genus region, along with a chosen path into the higher-genus region.

Taking a family of such reference fibers and paths along the critical circle of an SPWF yields a one-parameter family of such diagrams, and this paragraph describes how such a family evolves when a reference path passes a cusp (it is

helpful to imagine the family of horizontal arcs in Figure 1b). As the family of paths approaches a cusp, there is a vanishing arc  $\nu$  whose ends correspond under  $\beta$  coming from the vanishing cycle just past the cusp. This arc shrinks to a point where  $p$  and  $p'$  momentarily meet, so that the diagram whose path intersects the cusp itself has only one distinguished point. Passing the cusp,  $p$  and  $p'$  separate again. Those vanishing cycles that intersected  $\nu$  form the new collection of vanishing arcs, while those ends that came from vanishing cycles disjoint from  $\nu$  (and were brought together with  $p$  and  $p'$ ) float away without re-separating. With this understood, it becomes clear that the surface diagrams coming from reference paths on either side of a cusp are identical up to isotopy.

We now describe a way to use these diagrams to relate the surface diagrams at either end of the deformation. Observing that the intersections can be canceled by an  $R_2$  move at any time during the deformation by Proposition 2, the tactic is to record the map in a way that allows one to deduce the surface diagram if such a premature cancellation were to occur. In this way, we will have essentially defined a one-parameter family of surface diagrams connecting the endpoints of the deformation. To begin, choose a reference point  $r_t(0)$  that remains in the lowest-genus region between the two  $R_2$  moves and remains in the highest-genus region otherwise. This reference point will be contained in a smooth embedded reference path  $r_t(s)$ ,  $s \in [-1, 1]$  that initially is contained in the highest-genus region, but between the  $R_2$  moves has its endpoints in the two highest-genus regions and connects the two intersection points by a path in the lowest-genus region. Certainly, the base diagrams as measured from various points in  $r_0$  are identical up to isotopy. As the initial  $R_2$  move occurs, the fiber over  $r_t(0)$  experiences surgery on the two disjoint vanishing cycles of the intersecting fold arcs, gaining two pairs of points as discussed in the previous paragraph. Each of these pairs should be counted with multiplicity two, because there are four total intersections of  $r_t$  with the fold locus. At this point, the vanishing cycles are also partitioned into three sets, one set for each endpoint of  $r_t$ , and a *third partition* consisting of those vanishing cycles whose fold arcs have moved entirely within one of the sides of the bigon (the two fold arcs that intersect each other at the initial  $R_2$  move, and so are adjacent to both of the highest-genus regions at the same time, have been turned into pairs of points). Decorated in this way, the fiber over  $r_t(0)$  gives a prescription for how reference fibers over the two highest-genus components are identified for each value of  $t$ .

**Definition 3.** Such a reference fiber, decorated with points and vanishing cycles according to some chosen reference path, will be called a *surgered surface diagram*.

When the critical circle is immersed, the following can occur: cusps can move into and out of the sides of the bigon and the vanishing cycles can move around within the fiber over  $r_t(0)$  by isotopies. When a cusp moves into one side of the bigon, the pair of points associated to its adjacent fold undergo the same kind of modification described above, joining and re-separating. When an entire fold arc along with its adjacent cusps becomes contained in one side of the bigon, its vanishing cycle enters the third partition. When the second  $R_2$  move occurs, these processes reverse: the third partition empties and the four pairs of points become identified, first to have two multiplicity-two pairs, then the points within these pairs become identified in surgery to obtain  $\Sigma_{g+1}$ . In this way it is possible to deduce the resulting surface diagram after a pair of  $R_2$  moves.

As for the isotopies mentioned in the beginning of the previous paragraph, the vanishing cycles and their endpoints are free to wander about within the fiber above  $r_t(0)$ , subject to the intersection condition and the condition that the vanishing cycles coming from the highest-genus regions are not allowed to cross any endpoints, because their fold arcs lie in the higher-genus fibers obtained by surgery on those points. This explains why the vanishing cycles of those fold arcs lying partially within the sides of the bigon are counted in the first or second partition: when the second  $R_2$  move occurs, such a vanishing cycle necessarily survives into the surface diagram as measured in the highest-genus region. The vanishing cycles in the third partition are free to cross over endpoints subject to the intersection condition, and it has been shown that such modifications are handleslides. Finally, it is also important to note that the pairs of points form a braid in the 3-manifold above the reference arc between the two intersection points at all times, so that any wandering performed by them corresponds to an isotopy of this braid. In particular, each point must take *the same path* back to where it began when it rejoins its counterpart in the other partition, so that the vanishing cycles in the first and second partitions are not re-identified in a different way than when the deformation began. With this understood, the vanishing cycles before and after our deformation are related by some sequence of handleslides, up to isotopy.  $\square$

In the previous argument, a handleslide occurred exactly when a point wandered across a vanishing cycle. It will be important later to use the term *slide arc* to denote the fold arc in  $\text{crit } \alpha_t$ , bounded by two cusp points, whose vanishing cycle was crossed by the point. Note that there is a unique pair of self-intersection points in the critical image bounding the critical arc that contributed the point and the critical arc that contributed the vanishing cycle, corresponding to four points in  $\text{crit } \alpha_t$ ; the slide arc is then necessarily situated between two of these points, which are not mapped to each other by  $\alpha$ .

**3.2. Stabilization.** This move is a lightly disguised generalization of a homotopy that first appeared in Figures 5 of [B2] and 11 of [L1]. An example of how it affects a surface diagram is shown in Figure 4. On the left of Figure 4 is a small neighborhood  $\nu p$  of a point  $p$  of a circle  $c$  in a surface diagram  $(\Sigma_g, \Gamma)$ . Stabilization results in a diagram  $(\Sigma_{g+1} = \Sigma_g \#_{\partial \nu p} T^2, \Gamma')$ , with  $\Gamma'$  obtained as follows. First note that, without loss of generality,  $\Gamma \setminus c$  is disjoint from what appears in Figure 4, and so it is unambiguous to say that the collection of circles  $\Gamma \setminus c$  is preserved under stabilization. Then  $c \setminus \nu p$  patches in smoothly with one of the lower curves in the right of Figure 4. The next curve in the sequence is either one of the parallel circles, followed by the circle at the very top of the torus. The sequence continues with the other one of the parallel circles, and finally a parallel copy of  $c$  patches

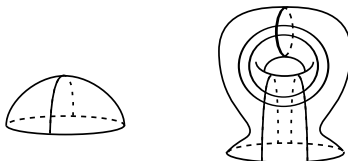


FIGURE 4. Stabilizing a surface diagram. The two middle arcs could be arbitrarily Dehn twisted about the top arc.

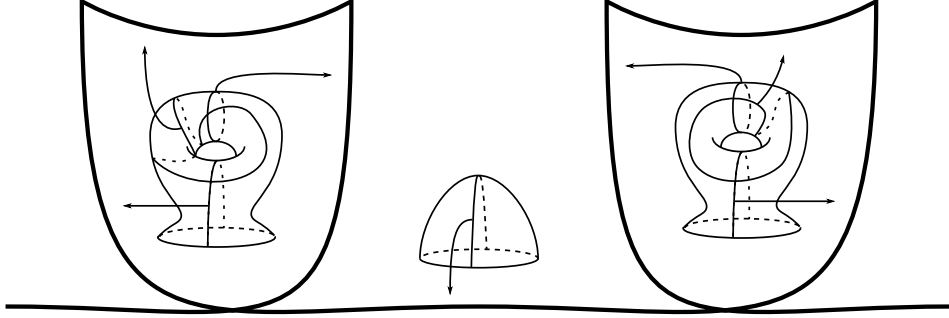


FIGURE 5. The stabilization deformation involves two flipping moves that occur on the same fold arc. The disk in the left side of Figure 2 is the middle fiber, and vanishing cycles are pictured as they might appear after performing the second  $R_2$  move to highlight the variety available in stabilizations.

in with the other curve at the bottom of the torus. The list of vanishing cycles then resumes as it did for  $\Gamma$ . Thus a *stabilization at  $p \in c$*  could be said to insert four curves into  $\Gamma$  along with a torus connect summand into  $\Sigma$ . Figure 5 depicts a base diagram for part of the stabilization deformation. Here, two *flipping moves* from [L1] have been applied to the same fold arc. For a single flip, there is a local parameterization of the fiber such that the vanishing cycles appear in a standard way as in Figure 5 of [L1] (the twice-punctured torus fiber in that figure is obtained from Figure 4 by attaching a two-dimensional one-handle containing the rest of the two parallel vanishing cycles not fully depicted). However, there is no reason to expect that either flipping should be required to appear in that standard way: following the constraints given by the intersection condition and the local model for the flipping move, the variety in how Figure 5 can appear is limited to allowing the vanishing cycles (call them  $v, v'$ ) for the fold arcs at the tops of the two loops to vary by positive or negative Dehn twists along the vanishing cycles for the central fold arc that runs between the two triangles; examples are shown in the figure. The stabilization deformation then continues as in [B2] and [L1], canceling the two intersection points by an  $R_2$  move (valid by Proposition 2) between the central fold arc and the critical arc connecting the two ends in Figure 5, resulting in a new simplified purely wrinkled fibration whose vanishing cycles can be deduced with little difficulty. To allow repeated stabilization, the  $R_2$  move requires the two-sphere to be the base of the fibration.

Another way to explain the variation of  $v$  and  $v'$  is to use surgered base diagrams as in the proof of Lemma 1. Using the reference point that gave the initial surface diagram, after performing the flips one may choose reference paths that enter the tops of the loops in Figure 5. After the  $R_2$  move, this gives a pair of surgered surface diagrams for  $M$  whose reference paths intersect a pair of fold arcs that are separated by one intervening fold arc (the central one in Figure 5). Bringing these two reference paths together to coincide with a reference path that crosses that intervening fold arc results in a pair of diffeomorphic surgered surface diagrams. As these diagrams are obtained from the same parametrization of the reference fiber, the diffeomorphism relating the two diagrams must be isotopic to the identity.

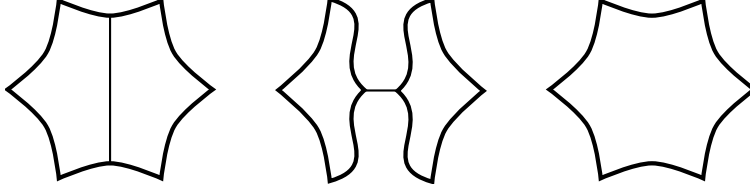


FIGURE 6. The multislide deformation. The initial fibrations for this and Figure 3 have six cusps for no particular reason.

The only such diffeomorphisms that could result in differing (non-surgered) surface diagrams are powers of Dehn twists along small circles around each of the pair of points coming from that fold arc, which is precisely what causes the variation in stabilizations, since  $v$  and  $v'$  are precisely those vanishing cycles affected by such diffeomorphisms.

One special case of stabilization is available to surface diagrams that come from surface bundles over the sphere, that is, blank surface diagrams. Performing a birth move in this case increases the genus of the surface diagram by one and introduces two simple closed curves whose only requirement is that they intersect at a unique point.

**3.3. Multislide.** Independently found by Denis Auroux and Rob Kirby, this move comes from a deformation in which the critical locus becomes momentarily disconnected. To perform the most complicated of the three moves, one first finds a nonconsecutive pair of vanishing cycles that intersect transversely at a unique point in  $\Sigma$ , say  $\gamma_1$  and  $\gamma_n$ . One also chooses one of the subsets  $\{\gamma_2, \dots, \gamma_{n-1}\}$  or  $\{\gamma_{n+1}, \dots, \gamma_k\}$  on which to perform the move: choose the first subset for expediency. A sufficiently small tubular neighborhood of the chosen pair  $\gamma_1, \gamma_n \subset \Sigma$  is a punctured torus, and one may replace that torus momentarily with a disk that has marked points on its boundary according to the places where  $\gamma_1, \dots, \gamma_k$  may have entered and exited the torus, resulting in a surface  $\Sigma'$  whose genus is one less than that of  $\Sigma$ , decorated with the resulting collection of curves  $\gamma'_1, \dots, \gamma'_k$  (there will be one pair of marked points for each intersection between  $\gamma_1, \dots, \gamma_k$  and the chosen pair). One may then perform a one-parameter family of diffeomorphisms of  $(\Sigma', \gamma'_2, \dots, \gamma'_{n-1})$  in which the disk travels along an arbitrary circle, returning to where it started and dragging the chosen subset  $\gamma'_2, \dots, \gamma'_{n-1}$  along with it (but leaving all the other vanishing cycles  $\gamma_n, \dots, \gamma_1$  unchanged). A choice of framing also allows those elements of  $\{\gamma'_2, \dots, \gamma'_{n-1}\}$  with marked points to be affected by a power of a positive or negative Dehn twist along the boundary of the disk. After performing this modification of the chosen subset, one reattaches the punctured torus according to the marked points. To explain the terminology, it is as if a handleslide has been performed in which a collection of vanishing cycles slides over the chosen pair. This is not to say the multislide is a sequence of handleslides: due to the nature of its deformation in which the cusps can trace out noncontractible circles in  $M$ , the multislide can change the free homotopy class of the underlying fibration's critical circle, something the previous two moves cannot achieve.

The multislide deformation, in the special case where the initial map has connected and embedded critical locus, is depicted in Figure 6. To begin, choose a pair

of fold arcs whose vanishing cycles happen to intersect transversely at a unique point in the fiber. In the left side of Figure 6, a vertical arc joins the chosen pair, signifying a move that results in the middle figure (this move is called a *merge* in [L1], and an *unmerge* in [GK1]; from habit this author tends to regress to the notation of [L1]). The inside of each of the two circles now has  $\Sigma$  as its generic fiber, decorated with vanishing cycles obtained from  $\Gamma$  by deleting those whose fold arcs were “pinched off” into the other circle by the merging move (this is where the partition  $\Gamma \setminus \{\gamma_1, \gamma_n\} = \{\gamma_2, \dots, \gamma_{n-1}\} \cup \{\gamma_{n+1}, \dots, \gamma_k\}$  comes from). Before any other Reidemeister or Lekili moves occur, the deformation concludes with an inverse merge in which the two circles reunite along the two newly-formed cusps (the two points at which the merge and inverse merge occur in such a deformation will be called a *multislide pair* even when the critical image may not be connected or embedded; c.f. Lemma 4). To explain the corresponding modification of surface diagrams, consider the deformation as a map  $[0, 1] \times M \rightarrow [0, 1] \times S^2$  that restricts to the identity map in the first factors. The critical image is a properly embedded twice-punctured torus in  $[0, 1] \times S^2$ . If one chooses two reference points in the left side of the figure such that one splits into each circle, initially their base diagrams are isotopic. However, the identification between them is modified by the time the two circles rejoin: note that the cusps that form and then disappear trace out a circle in  $[0, 1] \times M$ . Projecting out the homotopy parameter yields a circle  $\alpha \subset M$ , and without loss of generality one of the pair of cusps, say the right, remains stationary throughout its life, while the other traces out  $\alpha$  as the homotopy progresses (the resulting surface diagram will then be the one obtained using a reference point that stays in the right circle through the deformation). To say it a different way, the horizontal arc in the middle of Figure 6 can be taken as the image of that part of  $\alpha$  over which the left cusp will travel. This so-called *joining curve* is framed, is everywhere transverse to the fiber, and has the two cusps as its endpoints (at the moments when the cusps form and annihilate,  $\alpha$  appears as an actual circle in  $M$ ). The fibers containing points on its interior serve as  $\Sigma'$ , with the disk mentioned above coming from a fiberwise neighborhood of the joining curve. The circle  $\alpha$  projects to the circle in  $\Sigma'$  along which this disk travels, inducing the above-mentioned isotopy of the vanishing cycles in the left circle, which in turn induces an element of the mapping class group of  $\Sigma$  that only applies to those vanishing cycles unique to the left circle.

Choosing the left cusp instead of the right as the stationary cusp (and the left circle’s reference point as the ending reference point) results in performing exactly the reverse modification to  $\gamma_{n+1}, \dots, \gamma_k$ , moving along the circular path in the opposite direction and applying oppositely oriented Dehn twists along the boundary of the disk. For completeness, it should be mentioned that the surgered base diagram explanation for the multislide involves a reference path connecting the two newly-formed cusps throughout their existence; the details are left to the reader.

**3.4. Shift.** This move could be seen to embody the possible variations on the ordering of the vanishing cycles in a surface diagram, though it does not simply re-index the elements of  $\Gamma$ . In base diagrams, the deformation is similar to that of the multislide in Figure 6, but instead of reuniting the two circles at the two newly-formed cusps, one of those cusps instead meets with a cusp which is adjacent to the other. The two merge points at either end of a shift deformation will be called a *shift pair*.

Surgered surface diagrams offer a way to understand how this deformation affects a surface diagram. For the shift deformation, we choose a reference point in the lower-genus region and a fold arc  $\gamma$  that takes part in the initial merge. Also, choose two reference paths from that point that both cross  $\gamma$ , ending at points that will separate into the higher-genus side of each circle when the merge takes place. Initially this results in a surgered surface diagram whose points and vanishing cycles are all counted with multiplicity two, partitioned according to reference path. Once the merge takes place, each partition loses the decorations that were pinched off into the other circle. One of the reference paths crosses exactly one cusp in one of the circles to cross the fold arc that eventually joins the other fold arc crossed by the other reference path. The vanishing cycles of that partition change as described in the proof of Lemma 1, and when the circles reunite, each partition regains the elements of the other, with either one serving as the decorating set of the surgered base diagram.

It is important to note that the above modification works with following assumption. Once the initial merge occurs, there is a path traced out by one of the resulting cusps, with the other cusp remaining stationary in  $M$ . Slightly after the initial merge, the endpoint of this path (whose interior consists of regular points at all times) travels along exactly one fold arc to the cusp involved in the inverse merge, then serves as the joining curve. In other words, there is a disk in  $M_{[0,1]}$  consisting of joining curves between the cusps involved in the shift, except a small subinterval of its boundary circle  $c$  consists of fold points. This is unlike the multislide move, in which the analogous circle consists entirely of cusp points and may not bound a disk. Suppose we are given a deformation that corresponds to a shift, except that  $c$  projects to a circle  $c' \subset \{t\} \times M$  representing a nontrivial element of  $\pi_1(M)$ . We decompose this into a multislide and a shift as follows. Immediately after the initial merge, it is possible to inverse merge along a joining curve corresponding to  $c'$  and immediately re-merge along a joining curve corresponding to the inverse of  $c'$ . In the critical surface, it looks like a self-connect sum along a pair of short cusp arcs containing the initial merge point. In other words, we have performed a short one-parameter family of inverse merges. The result is a multislide followed by a shift.

#### 4. PROOF OF THEOREM 1

In what follows, the notation  $M_I$  for  $I \times M$  where  $I$  is an interval will be convenient, and maps  $M_I \rightarrow S_I$  will always be the identity on the first factor (that is, they are homotopies). As implied, the notation  $M_t$  will denote the slice  $(t, M)$ , and similarly for a deformation  $\alpha$ , the notation  $\alpha_t$  will refer to the map  $M_t \rightarrow S_t$ . Finally, we routinely conflate various maps with the fibration structures they induce. The criterion of a deformation  $f: M_{[0,1]} \rightarrow S_{[0,1]}^2$  to be generic implies the function  $T$  that projects  $\text{crit } f$  to the  $t$ -coordinate is Morse. Suppose  $\alpha$  is a homotopy between maps  $\alpha_i$  that induce surface diagrams  $(\Sigma_i, \Gamma_i)$ ,  $i = 0, 1$ . By the main result of [W] (and also [GK1] with the added value of a connected fibers result for deformations between maps with connected fibers), one may modify  $\alpha$  to be a *deformation of purely wrinkled fibrations* realized by applying a sequence of birth, merge, flip and isotopy moves to  $\alpha_0$ . For this reason, we shorten the terminology to simply *deformation* for such maps; these are the main focus of the paper. Examining the local models for birth, merge and flip, it becomes clear that the index zero and

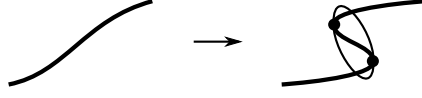


FIGURE 7. A pair of swallowtail points inserted into a cusp arc.

two critical points of  $T$  are precisely where births and their inverses occur, while the merging moves and their inverses occur precisely at the index one critical points of  $T$ . We routinely get away with calling these points birth and merge points, regardless of whether they correspond to births and merges or their inverses.

The proof Theorem 1 is broken into two parts: Section 4.2 is a series of lemmas to the effect that one may assume  $\text{crit } \alpha_t$  has at most two components at any value of  $t$ , and that  $T$  has no canceling Morse critical points. Viewing  $\alpha$  as a modification of  $(\Sigma_0, \Gamma_0)$ , this is the first step toward organizing  $\alpha$  into a sequence of model deformations, each corresponding to a move from Section 3; conceptually, these lemmas and those that follow make it so that multiple moves are not occurring at the same value of  $t$ . Section 4.3 involves a modification of the way the critical surface of such a deformation is immersed into  $S^2_{[0,1]}$ , essentially getting rid of the extraneous parts of the immersion locus that do not contribute to the modifications of  $(\Sigma_0, \Gamma_0)$ . What remains is the required deformation.

**4.1. Splicing cusp arcs.** This section gives a few tools which will be important for gaining some control over how the cusp locus is embedded in  $\text{crit } \alpha$ . The first is a local modification to introduce two swallowtail points into a cusp arc, as shown in Figure 7. In this figure, the two dots are the swallowtail points: the first one appears as a flipping move, while the other appears as an inverse flip. The bold arcs are those swept out by cusp points, while the circle on the right side is the locus of points at which the critical image of  $\alpha_t$  has self-intersections in  $S^2_t$  (see Section 4.3.1 for more about such depictions of the critical manifold of a deformation).

Figures 8 and 9 give base diagrams for the right side of Figure 7. At first, there is a cusp as in Figure 8a; a flip occurs, giving Figure 8b; the two upper vanishing cycles in the reference fiber follow directly from the local model for flips as described in [L1, W, GK1], while the lower two come from the fold arcs on either side of the original cusp. To obtain Figure 9a, the cusp at the lower left moves into the higher-genus region, with the vanishing cycles unchanged. Next, the cusp at the top left of Figure 9a moves into the lower-genus region resulting in Figure 9b, at which point the loop can be shrunk away by an inverse flipping move, and the upper two vanishing cycles survive.

It is also possible to cancel two swallowtails that are connected by an arc of fold points as in Figure 10. In this situation, the base diagrams for the left side have an inverse flip occurring, only for a flipping move to later occur on the same fold arc. Taking advantage of the local model for flips,

$$(2) \quad (t, x_1, x_2, x_3, x_4) \mapsto (t, x_1, x_2^4 + x_2^2 t + x_1 x_2 \pm x_3^2 \pm x_4^2)$$

in which a flipping move occurs at  $t = 0$ , the left side of the figure has a local model obtained by replacing  $t$  with  $t^2 + \epsilon$  for some small  $\epsilon < 0$ ; increasing  $\epsilon$  to a positive value gives the right side. The gluing between the two local models of flips implicit in this construction comes from a normal framing to the fold locus along the arc of fold points connecting the two swallowtail points.

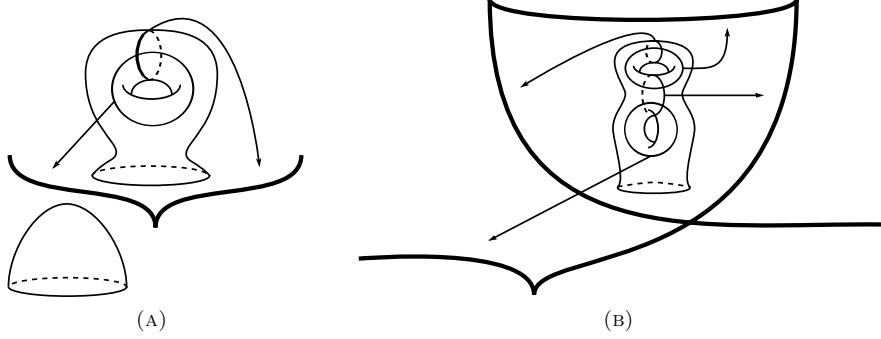


FIGURE 8. A pair of swallowtails introduced to a cusp arc, part one.

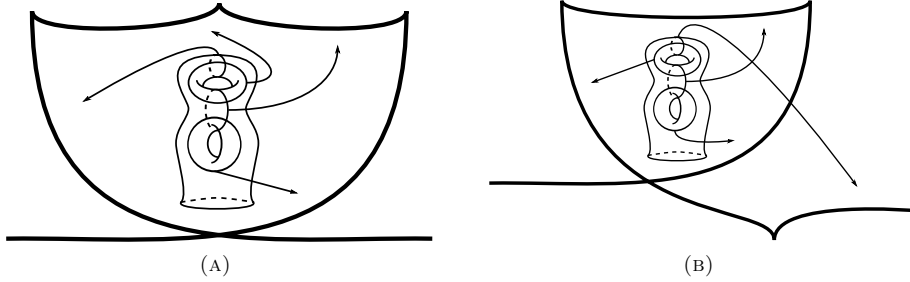


FIGURE 9. A pair of swallowtails introduced to a cusp arc, part two.

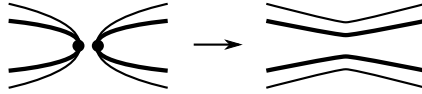


FIGURE 10. Canceling a pair of swallowtails.

Taking further advantage of the local model, it is possible to move a flipping move backward in  $t$ , or an inverse flipping move forward (within the same strip of fold points), by suitably extending the embedding of its model in that direction. This strings along a trail of cusp and immersion arcs in the wake of the swallowtail point, introducing pairs of cusp arcs and immersion points into the intervening maps  $\alpha_t$ .

Combining these moves yields a splicing modification reminiscent of the bypass move from contact topology, as in Figure 11. One introduces pairs of swallowtail points to three adjacent cusp arcs, then applies the canceling move twice. Applying this move inductively, it is possible to bypass across any odd number of cusp arcs along a reference arc that intersects them all either positively or all negatively.

#### 4.2. Modification of the critical manifold.



FIGURE 11. Bypassing an odd number of cusp arcs.

**Lemma 2.** Let  $\alpha$  be a deformation between simplified purely wrinkled fibrations. Then there is a deformation with the same endpoints whose critical locus is connected.

*Proof.* Denote by  $A_0$  the path component of  $\text{crit } \alpha$  whose image under  $T$  contains the minimum value of  $T$ , and suppose there is a path component  $A \subset \text{crit } \alpha$  distinct from  $A_0$ . Then  $A$  must have an index 0  $T$ -critical point because  $\text{crit } \alpha_0$  is connected, and by a homotopy of  $\alpha$  one may simply push that point backward in  $t$  until  $T(A) \cap T(A_0)$  contains an open interval, containing a generic value  $t_0$ . Let  $f: M_{[0, \epsilon]} \rightarrow S^2_{[0, \epsilon]}$  denote the deformation which is exactly  $\alpha_{t_0}$  for all  $t$ , except there appears the deformation appearing in Figure 10 of [L1] connecting arcs from  $A$  and  $A_0$ , then immediately the reverse of that deformation. The endpoints of  $f$  agree with  $\alpha_{t_0}$ , so that it is possible to insert  $f$  into  $\alpha$  at  $t_0$  to get a new deformation, resulting in the connect sum of the two components of  $\text{crit } \alpha$ . Now call the new map  $\alpha$  and denote  $A \# A_0$  by  $A_0$ . Repeating this process a finite number of times decreases the number of components of  $\text{crit } \alpha$  to one.  $\square$

**Lemma 3.** Let  $\alpha$  be the deformation resulting from Lemma 2. Then it is possible to remove all of the canceling critical points of  $T$ . More precisely, there is a deformation with the same endpoints satisfying exactly one of the following conditions:

- If  $\text{crit } \alpha_0$  and  $\text{crit } \alpha_1$  are nonempty, then  $\text{crit } T$  consists entirely of merge points.
- If exactly one of  $\text{crit } \alpha_0$  and  $\text{crit } \alpha_1$  is nonempty, then  $\text{crit } T$  has merge points and a single birth point.
- If  $\text{crit } \gamma_0 = \text{crit } \gamma_1 = \emptyset$ , then  $\alpha$  can be assumed to be the trivial deformation  $\alpha_t = \alpha_0$ ,  $t \in [0, 1]$ .

*Proof.* First assume  $\text{crit } \alpha_0$  and  $\text{crit } \alpha_1$  are nonempty, so that the goal, in other words, is to remove all the birth points. By Lemma 2,  $\text{crit } \alpha$  is connected so that for each index zero critical point  $b \in M_{t_b} \cap \text{crit } T$  (that is, for each birth point) there is an index one critical point  $m \in M_{t_m} \cap \text{crit } T$ ,  $t_b < t_m$  (corresponding to a merge or inverse merging move) such that  $b$  and  $m$  are a canceling pair of Morse critical points. Further, assume that  $m$  is the first such merge point with respect to  $t$  after possibly perturbing  $\alpha$  to ensure all merge points occur at different  $t$ -values. There are two subcases as follows.

In the first subcase, the cusp arcs containing  $m$  emanate forward in  $t$ ; that is, a merge occurs at  $m$ . Figure 12 depicts a special situation (falling under this subcase) in which there occurs a birth, creating two fold arcs  $a_1, a_2$  meeting at two cusps (appearing as concentric circles around  $b$ ). In base diagrams,  $a_1$  moves to intersect another nearby fold arc  $a_3$ , then  $a_2$  merges with  $a_3$ . The immersion locus appears as the two faintest arcs in the figure, and the two bolder cusp arcs break the surface into three pieces:  $a_1$  sweeps out the top piece,  $a_2$  the middle piece, and

$a_3$  the bottom. The result is as if the deformation actually consisted of two flipping moves, and in fact the two deformations are interchangeable within  $\alpha$ : one removes the part of  $\text{crit } \alpha$  corresponding to Figure 12 and instead glues in a flat disk of critical points containing two flips, with their accompanying cusp and immersion arcs. Any pair  $(b, m)$  in this subcase can be put into this situation by the following argument.

The cusp arcs and immersion arcs appearing in Figure 12 must also be present in  $\text{crit } \alpha$  because of the local models for merges and births. Further,  $\text{crit } T$  is limited to  $b$  and  $m$  in the figure (recall  $m$  is the first canceling index one critical point of  $T$  after  $b$ ), so that the differences between Figure 12 and what occurs in its counterpart in  $\text{crit } \alpha$  are limited to the addition of immersion arcs originating from  $R_2$  moves, and the addition of swallowtail points with their accompanying immersion and cusp arcs. Without loss of generality, each cusp or immersion arc limits to at most one swallowtail point by taking bites out of the edges of the picture.

If a swallowtail point is situated in the middle region, then its cusp arcs are contained in the middle region, too, because the cusp locus is embedded in the fold locus. Recall that a flipping move opens up a new arc of fold points bounded by two newly introduced cusp points. If the cusp arcs adjacent to the swallowtail point do not bound a disk in the middle region, the merge would necessarily be taking place between that newly introduced arc of fold points and  $a_3$ . This is impossible, because (choosing a suitable reference point in a base diagram) the vanishing cycle of that newly introduced arc is always disjoint in the fiber from all the preexisting vanishing cycles, other than the vanishing cycle of  $a_2$ , the fold arc on which the flipping move occurred. For this reason, all flipping moves result in cusp arcs that bound disks in their respective regions and immersion arcs that are typical of Reidemeister type one moves, so they can be pushed off the picture to the left or the right. As for immersion arcs, which serve as markers that prescribe the immersion of  $\text{crit } \alpha$  into the base of the deformation, it is evident that they may survive by appropriately including them into the deformation consisting of two flips, because of the agreement of vanishing cycles between the two deformations.

In the second subcase, the cusp arcs containing  $m$  emanate backward, corresponding to an inverse merge. Again, a special case of this is depicted in Figure 15a, in which a cusp arc coincides with the profile of the pictured disk of critical points.

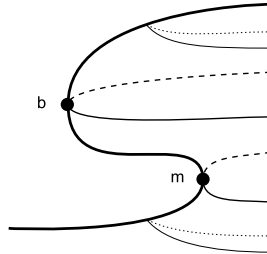


FIGURE 12. The critical locus for a birth followed by a merge, as one moves from left to right. This can be replaced by two flips according to Example 1 of [W] when it appears as above.

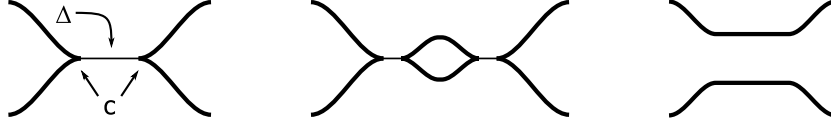


FIGURE 13. A perturbed version of the inverse merging move. Also, this is a cross-section of the surgery on crit  $\alpha$  in the proofs of Lemma 3 and the simply connected case of Lemma 4.

In base diagrams, a birth occurs near a cusp and later there occurs an inverse merge between that cusp and one of the cusps resulting from the birth.

As indicated by Figure 15b, one way to eliminate canceling Morse critical points on a surface is to take an appropriate connect sum with a sphere. In our situation, the way to achieve this is to find a smoothly embedded disk

$$\Delta: \{z \in \mathbb{C} : |z| \leq 1\} \rightarrow M_{[0,1]}$$

such that  $\Delta_t$  is a line segment for each  $t$  and such that the arc

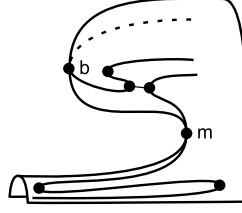
$$\Delta^{-1}(\text{crit } \alpha) = \{|z| = 1, \text{Re } z \geq 0\}$$

maps to a line segment  $\sigma$  that reaches slightly before  $t_b$  and travels to  $m$ , then backward to  $b$ .

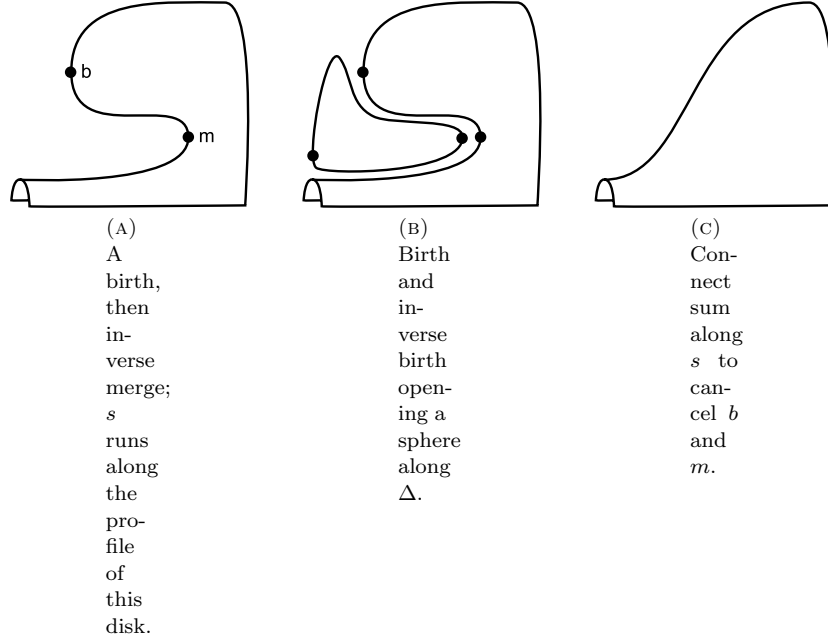
Supposing  $\sigma$  turned out to be a cusp arc, one could perform a birth move near  $\Delta(0)$  and immediately perform the inverse birth to remove the newly introduced circle of critical points, in effect opening a sphere  $S$  of critical points in  $M_{[0,1]}$  with a cusp equator. Expanding this sphere along  $\Delta$  until its equator coincides with  $\{\Delta(z) : |z| = 1 - \epsilon\}$  for small  $\epsilon > 0$ , one could then perform as in [S] a one-parameter family of inverse merging moves between  $S$  and crit  $\alpha$  along  $\sigma$ . A slight perturbation near the endpoints of  $\sigma$  would then remove crit  $T$  from the region.

A cusp arc coinciding with  $\sigma$  is easy to construct using the methods of Section 4.1. To restate the situation, a birth occurs and later there is an inverse merge between the resulting circle and the rest of the critical locus. By hypothesis, the only way for extra cusps to form on the birth circle is by the flipping move, so that there are exactly two possibilities: either  $b$  and  $m$  can lie on the same cusp arc (in which case no modification is necessary to connect them), or  $m$  lies on a cusp arc that reaches toward  $b$ , only to terminate at a swallowtail (in other words, a flip occurs on the birth circle and one of the resulting cusps inverse merges with the rest of the critical locus at  $m$ ). In the second case, it is straightforward to splice the cusp arcs containing  $b$  and  $m$  as depicted by the short arc connecting swallowtail points in Figure 14, using the cancellation trick as in Figure 10. Thus  $b$  and  $m$  are connected by a cusp arc, and it remains to extend the cusp arc on the other side of  $m$  backwards along the fold locus of  $\alpha$  to a point slightly before  $T_b$  as in the lower part of Figure 14. This completes the argument for removing birth points from  $\alpha$ , resulting in a new deformation; applying the same argument with  $t$  reversed removes the inverse birth points in a symmetric manner.

Two cases remain in the proof, as follows. If  $b$  minimizes  $T$ , then crit  $\alpha_0$  is empty and one simply ignores  $b$ , applying the previous argument to all other birth points, with  $b$  serving as the unique birth point coming from an initial stabilization of  $\alpha_0$ , creating a critical circle that survives through the deformation to be crit  $\alpha_1$  (one

FIGURE 14. Stringing a cusp arc along  $\sigma$ .

reverses the  $t$  parameter if  $\alpha_1$  is the side with empty critical locus). Finally, it may happen that  $\text{crit } \alpha_0 = \text{crit } \alpha_1 = \emptyset$ . In this case, the corresponding surface diagrams are both empty and  $\alpha$  can be assumed to be the trivial deformation,  $\alpha_t = \alpha_0$  for all  $t \in [0, 1]$ , because of the uniqueness of surface bundles of genus at least three over the two-sphere.  $\square$

FIGURE 15. Canceling critical points of  $T$  in the proof of Lemma 3.

**Lemma 4.** The deformation  $\alpha$  resulting from Lemma 3 can be modified so that every index one critical point of  $T$ :  $\text{crit } \alpha \rightarrow [0, 1]$  is part of a shift or multislide pair. If  $M$  is simply connected, the multislide pairs can be omitted.

*Proof.* Denote the first merge point at  $t = t_1$  by  $m_1 \in \text{crit } \alpha_{t_1}$ . Because  $\text{crit } \alpha_0$  is connected,  $\text{crit } \alpha_t$  has two components for  $t$  slightly larger than  $t_1$ . Choose two smooth curves  $c_{[t_1, t_2]}^i \rightarrow \text{crit } \alpha_{[t_1, t_2]}$ ,  $i = 1, 2$ , satisfying the following conditions:

- $c_{t_j}^i = m_j$ , where  $i, j \in \{1, 2\}$ .
- $c_t^1$  and  $c_t^2$  lie in separate path components of  $\text{crit } \alpha_t$  for  $t \in (t_1, t_2)$ .

The above conditions imply that the point  $m_2$  is the first merge point at which the path components of  $\text{crit } \alpha_t$  containing  $c_t^i$  become reunited. There exists such  $m_2$  because  $\text{crit } \alpha_1$  is connected. The idea of the proof is to first use the splicing deformation to introduce a cusp arc which is in some sense as isotopic as possible to the circle  $c = c^1 \cup c^2$ , then use it in a surgery to either shrink  $c$  to size or, in some cases, surger it out completely, decreasing the genus of  $\text{crit } \alpha$  by 1. The result then follows inductively.

For  $m_1, m_2$  to form a shift or multislide pair, it is necessary for a merging move to occur at  $m_1$  and an inverse merge at  $m_2$ . If this is not the case for some  $m_i$  we call  $m_i$  *problematic*. To remedy this issue we use a trick originally due to Auroux, enabling one to switch between a merge and inverse merge in the presence of a flipping move, as follows.

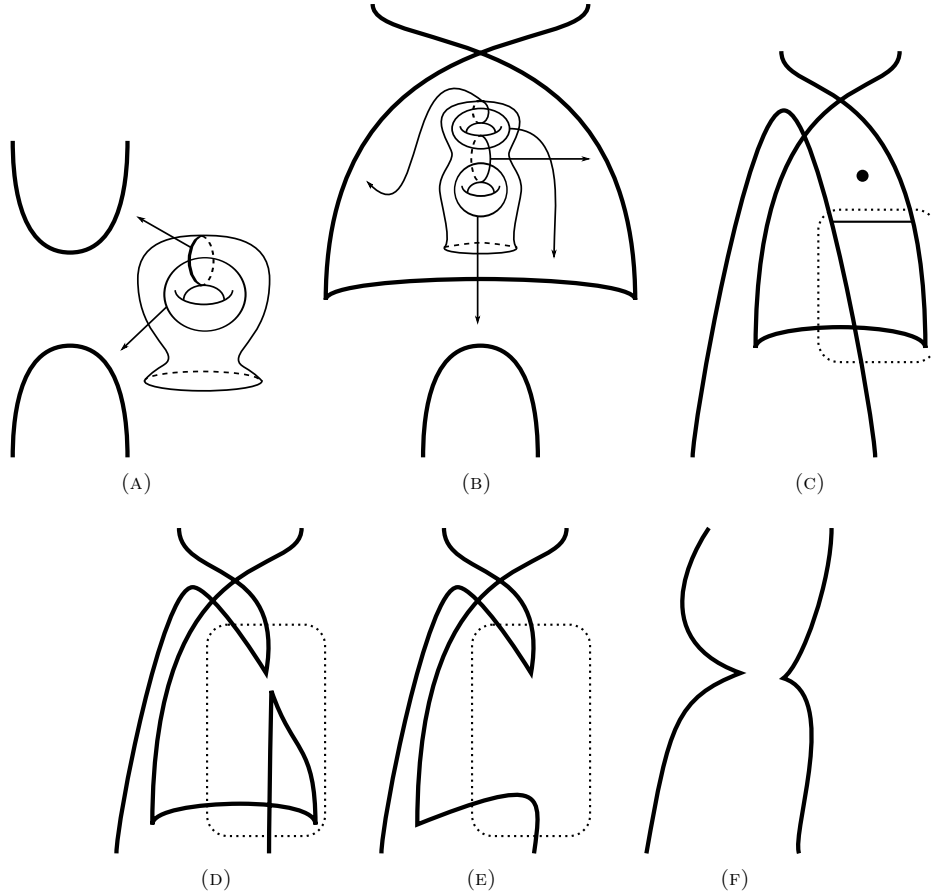


FIGURE 16. Putting a merge point into position for reversal.

If  $m_1$  is problematic, the first step is to put  $m_1$  into a situation to which Auroux's trick applies (if  $m_2$  is problematic, the following modifications all apply with  $t$

reversed). For this one may introduce a pair of swallowtails along its cusp arc qualitatively like in Figure 7, except the two swallowtails lie before and after the merge point with respect to  $t$ ; the precise modification appears in Figure 16. The way to interpret this figure is as a replacement for the merge deformation, whose base diagrams are given by Figure 16a followed immediately by Figure 16f; the validity of the replacement follows from the validity of the intervening base diagrams and the fact that the modifications therein occur relative to the fibration above the boundary of the target disk. The modification proceeds with a flipping move followed by an  $R_2$  move to obtain Figure 16c; the vanishing cycles follow directly from the local model for the flip and the  $R_2$  move is valid by disjointness of vanishing cycles. The decorated reference fiber in Figure 16b can be transferred to be the reference fiber above the dot in Figure 16c, with arrows pointing at the same fold arcs as before. In other words, the fold arcs of Figure 16c *inherit* their vanishing cycles from Figure 16b and we will use that term repeatedly in such contexts. Performing the indicated merge to obtain Figure 16d, the next steps are to cancel the top two intersections with an  $R_2$  move (which is valid by Proposition 1) and to close off the lower loop by an inverse flip, which can be seen to be valid by drawing in the three relevant vanishing cycles inherited from those of Figure 16b. The deformation above the three dotted rectangles now contains a merging move that can be replaced by one containing an inverse merge using Auroux’s trick, which follows.

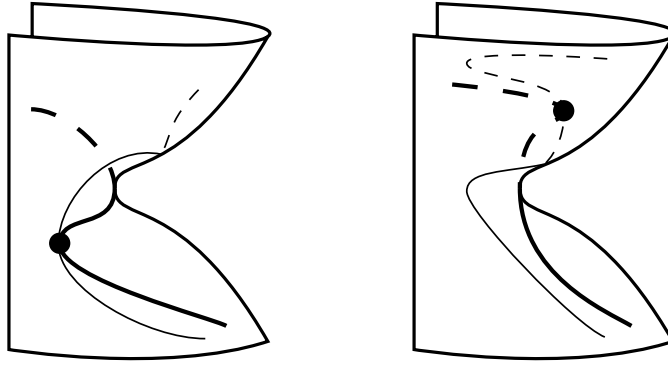


FIGURE 17. Sliding a swallowtail over a merge point can switch between merge and inverse merge.

The left side of Figure 17 is a depiction of the saddle-shaped critical surface in  $M_{[0,1]}$  above the dotted rectangles in Figure 16 (for  $t$  increasing left to right, it corresponds to the reversed order 16e, 16d, 16c because it is easier to depict and validate the deformations this way). As usual, the bold arcs are cusp arcs, and in each figure  $m_1$  lies at the saddle point of the surface, with a flipping move occurring at the dot. The fainter arcs are where  $\alpha$  is two-to-one on its critical locus; for more details on such pictures, see Section 4.3.1. It is not difficult to deduce the base diagrams specified by the left side of Figure 17 by considering vertical slices of the figure from left to right: the progression for the left side appears in Figure 18a, which is a copy of what happened in the dotted rectangles from before. Here a fold

arc experiences a flip, then one of the resulting cusps experiences an inverse merge with another preexisting cusp.

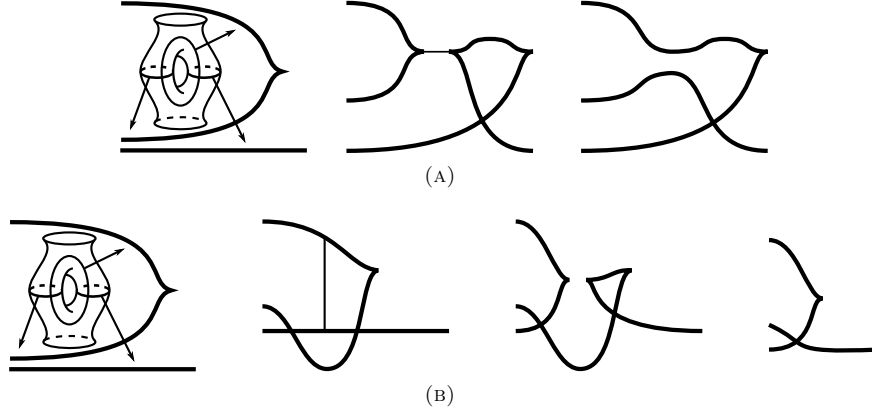


FIGURE 18. Base diagrams for the left and right side of Figure 17, respectively.

Obtained by an isotopy of  $\alpha$  that involves sliding the swallowtail point along the cusp arc containing it, the right side of Figure 17 has base diagrams given by Figure 18b, which will replace those of Figure 18a in the dotted rectangles. Again, following vertical slices from left to right, one may deduce that the sequence of moves must begin with an  $R_2$  move between a fold arc and another fold arc containing a cusp. The deformation concludes with a merging move and an inverse flip as appears in Figure 18b. The validity of these moves and the intended substitution follow entirely from the vanishing cycles in the initial base diagrams in Figure 18, which are inherited from the base diagrams of Figure 16. In Figure 18b the initial  $R_2$  move is valid because the vanishing cycles involved are disjoint. The following merge move and inverse flip are straightforward to verify using the vanishing cycles inherited from the initial base diagram. Restating the prescription, one first substitutes the reverse of Figure 18b into the dotted rectangles of Figure 16, and then substitutes the resulting deformation for a neighborhood of a problematic merge point  $m_1$ . Thus is safe to assume that  $m_1$  and  $m_2$  are not problematic.

The path  $c^i$  generically lies in the fold locus of  $\alpha$ , and intersects the cusp locus transversely at  $k$  points, including  $m_1$  and  $m_2$  (these are cusp points by Proposition 3 of [W]). If  $k$  is odd, the splicing and extending methods of Section 4.1 allow one to arrange that  $m_1$  and  $m_2$  lie within the same cusp arc. If  $k$  is even, it is possible to arrange for the cusp arcs containing  $m_1$  and  $m_2$  to be adjacent in the part of  $\text{crit } \alpha_{(t_1, t_2)}$  containing  $c^i$  (that is, there is an arc of fold points in  $M_t$  connecting them for each  $t \in [t_1, t_2]$ ).

After applying these modifications to  $c^1$  and  $c^2$ , the next step is to use something similar to the modification that appeared in Figure 15: choosing small  $\epsilon > 0$ , perform two births at  $t = t_1 + \epsilon/2$  whose circles come together by inverse merge at  $t = t_1 + \epsilon$ . The resulting circle survives until an inverse birth at  $t = t_2 - \epsilon$ , giving a cusp arc (bounded by the two birth points) that approximates the one containing  $m_2$ . Performing inverse merges along this cusp arc moves  $m_2$  to  $t_1 + \epsilon$ . If  $m_1$  and

$m_2$  lie on a single cusp circle, then they now form a multislide pair for  $\epsilon$  sufficiently small. If  $M$  is simply connected, then this cusp circle bounds a disk

$$\Delta: \{z \in \mathbb{C} : |z| \leq 1\} \rightarrow M_{[t_1, t_1 + \epsilon]}$$

such that  $\Delta_t$  is a joining curve in  $M_t$  for  $t \in (0, 1)$ , allowing one to replace  $c$  with a pair of disks. For this reason, all multislide pairs can be eliminated if  $M$  is simply connected.

For shift pairs,  $m_1$  and  $m_2$  are required to be connected by a cusp arc isotopic in  $\text{crit } \alpha$  to  $c^1$  or  $c^2$ . That is, for a shift pair, a merge occurs, forming two cusps  $\chi_1$  and  $\chi_2$ ; immediately there occurs an inverse merge between  $\chi_1$  and a cusp  $\chi'_2$  adjacent to  $\chi_2$ . If  $m_1$  and  $m_2$  are not connected by a cusp arc isotopic in  $\text{crit } \alpha$  to  $c^1$  or  $c^2$ , then we have arranged for the merge to occur between cusps  $\chi'_1$  and  $\chi'_2$  which are adjacent to  $\chi_1$  and  $\chi_2$ , respectively. This is easily remedied by performing an inverse merge between  $\chi_1$  and  $\chi'_2$  at some  $t \in (t_1, t_1 + \epsilon)$  and immediately reversing it, breaking  $m_1, m_2$  into two shift pairs. Choosing  $\epsilon$  small enough, the above modifications may be repeated inductively, starting with the first merge point after  $t_1 + \epsilon$ .  $\square$

**4.3. Modification of the immersion of the critical manifold.** At this point,  $\alpha$  has some of the characteristics of the required deformation: the Morse function  $T: \text{crit } \alpha \rightarrow [0, 1]$  and the cusp locus containing its critical points are as required; however, the base diagrams can still appear disorganized because of self-intersections in the critical image. The object of this section is to modify this immersion  $\text{crit } \alpha \rightarrow S^2_{[0, 1]}$  until the remaining double points all come from the model deformations for the moves in Theorem 1.

**4.3.1. Immersion loci of deformations.** It will be necessary to collect some facts before beginning the main argument of this section. To summarize the results of Section 4.2,  $\text{crit } \alpha_t$  is a single immersed circle that, as  $t$  progresses, can momentarily split in two at shift or (in the case that  $\pi_1(M)$  is nontrivial) multislide pairs.

The critical manifold of  $\alpha$  has various decorations: the immersion locus  $\iota$  consisting of paired immersion arcs, the cusp locus  $\chi$  which is a smoothly embedded one-submanifold of  $\text{crit } \alpha$ , and swallowtail points coinciding with  $\bar{\chi} \setminus \chi$ . As a final decoration, away from a tubular neighborhood of  $\iota$  one could indeed view  $\text{crit } \alpha$  as a kind of base diagram whose fiber above any point  $p \in \text{crit } \alpha_t$  is defined by  $\alpha^{-1} \circ \widetilde{\alpha(p)}$ , where  $\widetilde{\alpha(p)}$  is obtained by slightly perturbing  $\alpha(p)$  in  $S^2_t$  to a regular value in its higher-genus side (one may then extend to  $\text{crit } \alpha \setminus \iota$  by continuity). This defines a closed, orientable surface along with a distinguished simple closed curve, which is the vanishing cycle for  $\alpha(p)$  in a base diagram for  $\alpha_t$ . In this light, each immersion arc gains a vanishing cycle from the fold that contains its counterpart, and crossing an immersion arc has the same effect on the fiber as crossing a fold arc in a base diagram. Following fibers across a cusp arc, the distinguished vanishing cycle changes into another that transversely intersects the previous at a unique point.

We will call such a diagram a *critical base diagram*. Importantly, the one-parameter family of base diagrams for  $\alpha$  can be recovered from the critical base diagram by first using the immersion locus and merge points to recover the sequence of Reidemeister moves and merges in the critical image (the Reidemeister

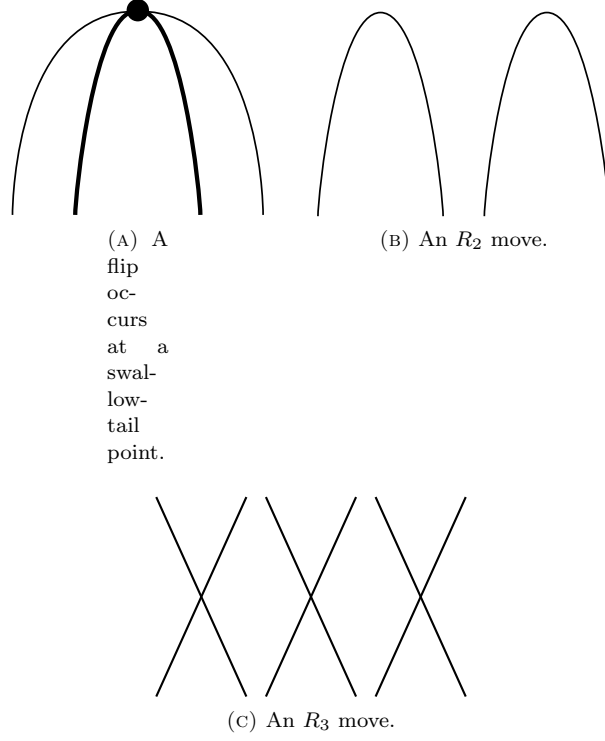


FIGURE 19. Critical and immersion loci of  $\alpha$ . Here the  $t$  parameter ranges vertically, cusp arcs are bold, and pairs of fainter arcs are mapped to each other by  $\alpha$ .

moves, limited to flips,  $R_2$  and  $R_3$  moves, are unambiguously specified by examining whether the genus increases or decreases as one crosses the relevant immersion arcs). After adding cusp points, one can then add regular fibers marked with vanishing cycles. Thus a sufficiently well-decorated critical base diagram is simply another way to depict a deformation. It is important to note that a base diagram does not in general specify the total space up to diffeomorphism; see for example the discussion around Figure 8 of [GK2]. In particular, critical base diagrams are a slightly lossy tool for recording a deformation, though they turn out to be sufficient for proving Theorem 1.

It follows from Lemma 1 of [W] that, as  $t$  increases or decreases, there are only two ways for self-intersections to form in the critical image of a deformation: flips and  $R_2$  moves ( $R_2$  moves for short). In a critical base diagram, a flipping move is encoded by a pair of immersion arcs nested around a pair of cusp arcs whose common endpoint is a swallowtail point as in Figure 19a. The  $R_2$  move appears as two immersion arcs  $\iota_1$  and  $\iota_2$  that appear as in Figure 19b.

An important feature of deformations, which will be called condition  $s_2$ , is that the maps  $T|_{\iota_i}$  each have a single common critical value for each  $R_2$  move; in other words, the arcs  $\iota_i$  are tangent to the same slice  $M_t$ , and such tangencies always come in pairs, one pair for each  $R_2$  move. Finally, there may be Reidemeister type

three moves, or  $R_3$  moves for short, whose immersion loci appear in  $\text{crit } \alpha$  as in Figure 19c. These also come with a simultaneity condition which will be called property  $s_3$ : the three intersection points between immersion arcs in the critical surface must all have the same  $t$ -coordinate. Since  $\alpha_0$  and  $\alpha_1$  are injective on their critical circles, the closure of the immersion locus is a union of circles; let  $c$  denote one of these (to avoid clutter in the exposition, it is convenient to include swallowtail points in the immersion locus).

**Lemma 5.** The following are equivalent for deformations whose critical locus is connected at each value of  $t$ .

- (1)  $c$  is one of a pair of circles which are mapped to each other by  $\alpha$
- (2)  $c$  bounds a disk  $D \subset \text{crit } \alpha$  and is free of swallowtail points.

*Proof.* (1)  $\Rightarrow$  (2): if  $c$  contained a swallowtail point, one could choose a generic point on  $c$  and mark its counterpart on the other circle. One could then follow the one-parameter family of double points along the immersion arcs until one of them enters a neighborhood of the swallowtail point as in Figure 19a, and here we come across a one-parameter family of triple points (or higher), the third arc coming from the other side of the immersion arc passing through the swallowtail point, which cannot be simplified by a small perturbation. This is a contradiction because in a deformation the only triple points are those coming from Reidemeister type three moves, which result in isolated triple points. Thus,  $c$  is free of swallowtail points. Now suppose  $c$  does not bound a disk in  $\text{crit } \alpha$ . Property  $s_2$  would then imply that  $c$  must transversely intersect its counterpart, and tracing identifications shows these intersection points must lie in the same slice  $M_t$ . This along with property  $s_2$  implies the other circle also must not bound a disk in  $\text{crit } \alpha$ ; in this case, each pair of intersection points between  $c$  and its counterpart can be resolved by an arbitrarily small perturbation of  $\alpha$ , resulting in two circles that bound disks in  $\text{crit } \alpha$  (think of performing an  $R_2$  move and momentarily pinching together the two middle arcs bounded by intersection points: the perturbation removes the momentary tangential contact).

(2)  $\Rightarrow$  (1): By continuity and the fact that generic immersion points of curves into surfaces are double points, if  $c$  is not one of a pair of circles then it is a single circle mapped by  $\alpha$  in a two-to-one fashion outside of a finite subset of points, each corresponding to a Reidemeister type one move, which necessarily occurs at a swallowtail point. Arbitrarily choosing an orientation of  $c$ ,  $\frac{\partial}{\partial t}$  projects to oppositely oriented tangent vectors at a chosen pair of points of  $c$  with a common image. A swallowtail may be located by tracing the pair (as a one-parameter family of double points) along  $c$  in the direction specified by the two tangent vectors, so that the two points must eventually coincide at a swallowtail point.  $\square$

**Definition 4.** Lemma 5 suggests the terminology *immersion pair* for a pair of circles in the critical locus which have the same image, and *immersion single* for a circle of immersion points containing at least one swallowtail point.

**Lemma 6.** One may perform  $R_3$  moves and cancel intersections by  $R_2$  moves performed on the immersion locus of a critical base diagram.

*Proof.* The three immersion arcs that form a candidate for an  $R_3$  move break the fold locus into a central triangle and, in a neighborhood of that triangle, six adjacent regions. The first step is to show that one may assume the central triangle is free

of cusp arcs so that the  $R_3$  move involves only the immersion locus. Toward this end, note that the assumption is that we are working with a triangle bounded by three immersion arcs, so that there are no swallowtail points in the central triangle by hypothesis. This leaves cusp arcs that wander in and then out of the central triangle, leading to a pair of intersection points between each and the boundary of the triangle. Along with cusp arcs come the issue of slide arcs. At most two sides of the triangle contain the pair of immersion points bounding the critical arc containing the slide arc, so that the third is allowed to pass over the slide arc, removing it from the triangle. It remains to remove cusp arcs from the triangle; the two intersections between the cusp arc and the boundary of the triangle may lie on one or two sides. If both lie on one side, one may perform an  $R_2$  move between the cusp arc and the immersion arc, removing it from the triangle: in base diagrams, the appearance is simply a cusp wandering along its fold arc across an intersection and back (the  $R_2$  move in the critical base diagram results in the cusp remaining on one side of the intersection). When the cusp arc enters at one side of the triangle and exits through another, what happens in base diagrams (possibly reversing  $t$  or even looking “sideways” as in the splicing deformation) is as follows. Coming from the first corner of the triangle, there is an  $R_3$  move between the images of three critical arcs, one of which contains a cusp. This cusp starts outside the relevant triangle, and once the  $R_3$  has occurred it moves along its fold arc all the way across one side of the newly-formed triangle to the other part of the fold arc not on the triangle. The observation here is that the round vanishing cycles on either side of the cusp arc and either side of each immersion arc are disjoint because of the way they intersect, so that the cusp could just as well have moved over before the  $R_3$  move. In the critical base diagram, the effect is an  $R_3$  move between the cusp arc and the immersion arcs it intersects, removing it from the triangle. Thus there is a neighborhood of the triangle of immersion arcs which may be assumed empty of all decorations.

There are four choices for the orientations of the triangle’s three immersion arcs:  $k$  of their three vanishing cycles can be measured using a reference point in the central triangle for  $k \in \{0, 1, 2, 3\}$ . For each  $k \neq 0$  it is straightforward to check that there is at least one “highest genus” region, in the sense that the fibers in the other regions are obtained from the fiber above a “universal” reference point  $p$  in that region by surgery on  $n \leq 2$  circles (in other words,  $p$  can serve as a reference point for finding the vanishing cycles for all three immersion arcs). As mentioned before, the three vanishing cycles as measured from  $p$  are pairwise disjoint so that an  $R_3$  move on the immersion arcs is valid. For  $k = 0$  the “lowest genus region” is the central triangle. In this case, the  $R_3$  move is valid because in this case one may choose a reference point opposite one of the corners of the triangle, and observe that two of the vanishing cycles can be seen to be disjoint in the fiber while the third immersion arc contributes a pair of points at which the reference fiber self connect sums; certainly the three surgeries corresponding to the pair of points and the vanishing cycles commute, so that the third arc can move past the chosen corner for an  $R_3$  move.

Now suppose two immersion arcs meet at two points in the manner typical of  $R_2$  moves. The argument that one may assume the bigon that constitutes an  $R_2$  candidate can be assumed free of cusp arcs and slide arcs is similar to that of the  $R_3$  move. There are three choices for the orientations of the two immersion arcs,

according to whether the universal reference point lies within the bigon, across one of its sides, or across one of its corners. In each case, the intersections between the immersion arcs requires their vanishing cycles to be disjoint, so that the  $R_2$  move is valid.  $\square$

#### 4.3.2. Unlinking $\iota$ .

**Lemma 7.** For the deformation  $\alpha$  as described at the beginning of Section 4.3.1, there is a modification that causes  $\alpha_t$  to be injective on its critical locus at those values of  $t$  for which  $\text{crit } \alpha_t$  has two components.

*Proof.* Another way to state the result is to say one can move  $\iota$  into the cylinders between multislide and shift pairs in the critical base diagram of  $\alpha$ . Suppose the first pair whose intervening pair of critical cylinders are immersed lies at  $\{t_0, t_0 + \epsilon\}$ . The issue we must address is that  $\iota$  may have *linked immersion circles*: these are immersion circles which are not freely homotopic to ones that lie outside of  $M_{[t_0, t_0 + \epsilon]}$ . These come in two types: *split immersion circles* (those linked components of  $\iota$  that intersect both components of  $\text{crit } \alpha_t$  for  $t \in [t_0, t_0 + \epsilon]$ ) and those that do not split. The three possibilities for a split immersion circle are shown in Figure 20, in which there could be an immersion pair with one or two split components or a split immersion single. The idea will be to un-split these circles in a such a way that an application of Lemma 6 allows an isotopy of  $\alpha$  that slides them off  $\text{crit } \alpha_{[t_0, t_0 + \epsilon]}$ .

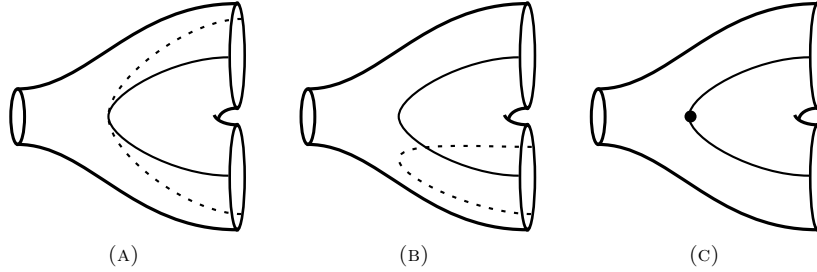
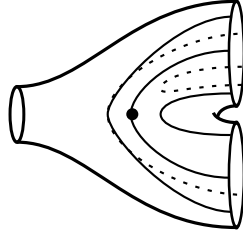


FIGURE 20. Various possibilities for the immersion locus near  $\text{crit } \alpha_{t_0}$ . Figures (b) and (c) are impossible unless there are other immersion arcs present because two circles in  $S^2$  generically intersect in an even number of points.

Using Lemma 6, nonlinked immersion pairs and singles can be retracted forward or backward until they are disjoint from  $\text{crit } \alpha_{[t_0, t_0 + \epsilon]}$ . By a further isotopy, push any  $R_3$  moves involving split circles forward in  $t$  until just after the first merging move occurs at  $t_0$ . This will arrange them so that they are nested approaching  $t = t_0$  as in Figure 21; this gives them an ordering and the lemma follows inductively by unsplitting the first split circle of  $\alpha$ , then sliding it off  $\text{crit } \alpha_{[t_0, t_0 + \epsilon]}$ .

As one travels across an immersion arc with increasing  $t$ , the Euler characteristic of the fiber increases or decreases by 2. If it decreases at both that arc and its counterpart, we call them *genus-increasing*. If it increases at both, it is *genus-decreasing*, and if it increases at one but decreases at the other, we call them *mixed*. The comments of the previous paragraph imply  $\text{crit } \alpha_t$  can be made embedded prior to the sequence of flips and  $R_2$  moves whose immersion arcs split at  $t_0$ . For this

FIGURE 21. Nested split immersion circles in crit  $\alpha$ .

reason, it is straightforward to verify that if the first split immersion arc is an immersion pair, then it is not mixed (going further, the fact that a merge always occurs on the higher-genus side of each of two fold arcs implies there can be no such thing as a mixed split immersion pair). This reduces the argument to three cases for linked circles: the genus-increasing and genus-decreasing pairs, and immersion singles. In each case, there is a split version and a nonsplit version, which have essentially the same remedy. Note that, in what follows, the modifications occur only near  $t = t_0$ , so that it is possible that some “pairs” eventually connect at some  $t > t_0 + \epsilon$ , and thus (without affecting the validity of the argument) turn out to be singles in the larger deformation.

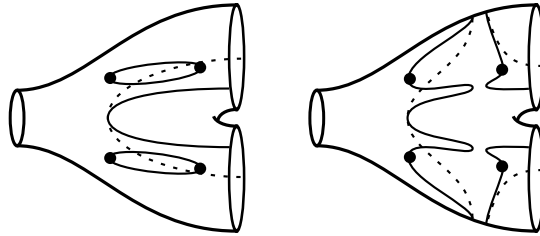


FIGURE 22. Critical base diagrams for un-linking Figure 20a in the genus-increasing case.

Case 1: genus increasing pairs. Suppose the initial split immersion arc belongs to a genus-increasing split immersion pair. In other words, the first split circle appears as in Figure 20a and passing across one of the circles causes the Euler characteristic of the fiber to decrease. In Figure 22, the modification to  $\alpha$  appears in a critical base diagram in which we introduce immersion singles by flipping then inverse flipping to get the left side, then “pinching” pairs of immersion arcs by an isotopy of  $\alpha$  to get the right side. In this figure, the dots are swallowtail points and cusp arcs have been omitted. After applying such a modification, the immersion arcs no longer split; it remains to show the validity of the figures. To this end, base diagrams for the right hand side of Figure 22 appear in Figure 23.

In Figure 23, cusps have been sprinkled about in order to give the reader some perspective on how such a deformation would appear; indeed, these modifications are valid regardless of the placement of cusps: what matters is the placement of the swallowtails in relation to the immersion locus. Figures 23a and 23b depict

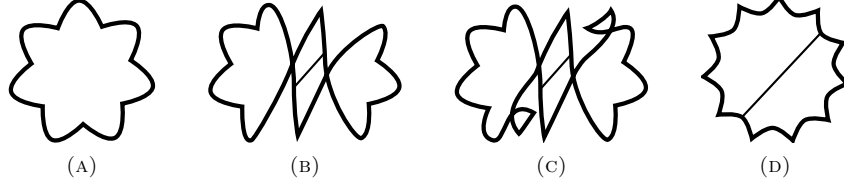


FIGURE 23. Base diagrams for the right side of Figure 22. The deformation proceeds from left to right, then back to the left before the indicated merge.

the initial hypothetical  $R_2$  move that produces the offending immersion pair. Immediately, as depicted in Figure 23c, two flips occur, and the isotopy to obtain Figure 23d is a pair of  $R_2$  moves, valid by Proposition 2. The deformation then reverses itself back to Figure 23a and  $\alpha$  resumes as it did before. Deducing the correspondence between Figures 22 and 23 is a moderately straightforward exercise. For a nonsplit pair (or any other case that arises from a genus-increasing  $R_2$  move), the same trick converts the two arcs that intersect  $[t_0, t_0 + \epsilon]$  into a pair of immersion singles, reducing to case 3 below.

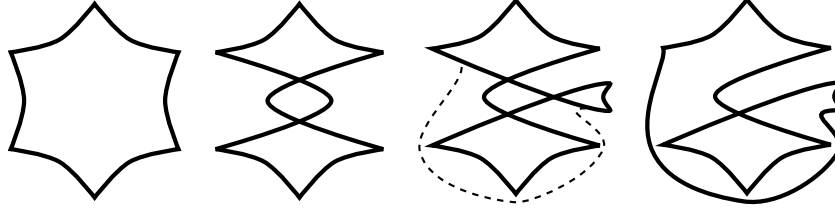


FIGURE 24. Converting a genus-decreasing split immersion pair into a pair of genus-decreasing split singles.

Case 2: genus-decreasing pairs. Suppose the initial split immersion arc belongs to a genus-decreasing split immersion pair. The same idea of canceling the double points formed by the  $R_2$  move against those formed by flipping moves still applies, though instead of unlinking the circles, it converts them into a pair of linked singles, reducing to the last case covered below. Figure 24 gives base diagrams for most of a deformation in which a flip occurs immediately after the initial  $R_2$  move that formed the offending immersion pair. This is followed by an isotopy as indicated by the dotted line. In this isotopy, the initial  $R_3$  move is valid by Proposition 3. The following  $R_2$  move can be validated by Proposition 1 after moving the cusps into the higher-genus side of the uncusped arc and possibly stabilizing at a fold arc in the stationary set of the shift or multislide, then de-stabilizing afterward, to satisfy the genus requirement. The same arc then continues around the back of the sphere to result in a base diagram with a figure-eight configuration; one may then perform a flipping move on the wandering arc from the last two steps, and then an  $R_2$  move (Proposition 2) cancels all intersection points. The deformation then proceeds precisely in reverse until it reaches the second map of Figure 24, at which

point the deformation continues as it did before the interlude above. Figure 25 gives the critical base diagram for this deformation and it is slightly more tricky (though still elementary) to deduce the correspondence between the figures. As this modification converts any genus-decreasing immersion pair into two immersion singles, the various ways of linking are all reduced to case 3.

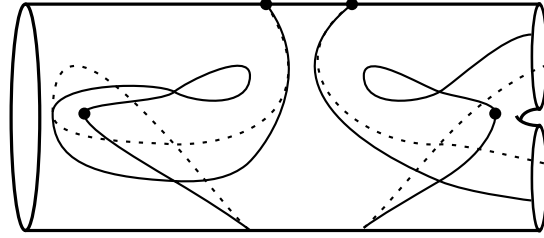


FIGURE 25. Critical base diagram for Figure 24. The merge point is included to give perspective on how the resulting immersion singles are linked.

Case 3: linked singles. Like in the case of immersion pairs, the first linked single could be genus-increasing or genus-decreasing. The two cases have a similar remedy; first come the details for nonsplit genus-decreasing singles. Figure 26 gives a standard way such things appear (a critical base diagram for this deformation, and one variation, which by Lemma 6 can be made standard, appear in Figure 25). Figure 27 describes a way, beginning with the right side of Figure 26, to move the two immersion arcs of a linked single from one cylinder in  $\text{crit } \alpha_{(t_0, t_0+\epsilon)}$  to the other.

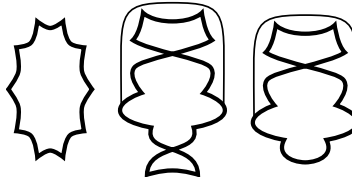


FIGURE 26. Genus-decreasing nonsplit singles appear as above, possibly after an application of Lemma 6. The sequence is an  $R_2$  move followed by an inverse flip. The merge occurs later between the indicated fold arcs.

Figure 27 is meant to describe a deformation  $\alpha^s$ ,  $s \in [0, 1]$  of  $\alpha = \alpha^0$  by depicting successive slices  $\alpha_\tau^s$  for a fixed value  $\tau < t_0$  as  $s$  increases ( $\tau$  is a  $t$ -value just after the inverse flip in Figure 26). For each  $s$ ,  $\alpha^s$  before  $\tau$  begins with the left side of Figure 26 and proceeds to the relevant step in Figure 27. As  $s$  increases, the two immersion arcs move one at a time from one cylinder to the other. It is interesting to note that linked immersion circles appear at those values of  $s$  at which the merging arc in Figure 27 crosses fold arcs, but are gone by the time  $s = 1$ . Furthermore, this modification contains a remedy for the split case as well: the left side of Figure 27 should then be considered part of Figure 26 as the typical way split singles arise in

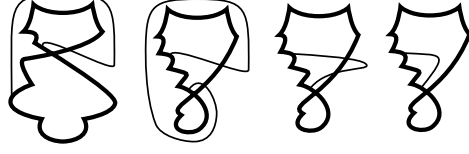


FIGURE 27. A one-parameter family of merge pairs where the parameter  $s$  increases to the right, described by a sequence of base diagrams. The merging arc sweeps around the back of the sphere between the second and third figures.

a deformation. After unlinking, the immersion arc moves forward in  $t$  past  $t_0 + \epsilon$  and the process continues with the next linked single.

For the genus-increasing case, after an isotopy of  $\alpha$  the immersion single initially appears by a swallowtail. If it splits, the merge occurs between a point in the “head” of the loop formed by the flipping move and a point away from the loop (see Figure 28). The former point cannot lie at the top of the head because that

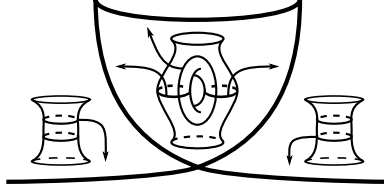


FIGURE 28. The loop that appears in the critical image immediately after a flipping move suggests coordinates consisting of *head* and *shoulders*.

vanishing cycle is always disjoint from all others due to the local model for the flipping move, so it must lie on one of the left or right sides of the head (the one exception is when the merge occurs between the top and one of the adjacent fold arcs; in this case momentarily closing off one of the critical components by an inverse birth, then opening it back up again with a birth, then applying Lemma 3 to get rid of the two birth points removes the multislide pair entirely). For this reason, the possibilities for genus-increasing linked immersion singles are limited to those whose base diagrams appear in Figure 29. In the split case the trick from the genus-decreasing case turns a split single into a nonsplit single, moving one of the ends of the merging arc (and some number of cusps) so that it is contained in the higher-genus region of the head, resulting in a nonsplit immersion single. The way to unlink in this case is similar to that of genus-increasing doubles: perform a flipping move on the lower part of Figure 29b and cancel the two intersections with an  $R_2$  move in a generalized stabilization, then immediately reverse back to the map of Figure 29b. The critical base diagram in Figure 30 illustrates the unlinking.  $\square$

4.3.3. *Organizing the unlinked immersion locus.* This section lists the last few modifications to  $\text{crit } \alpha$ . According to Section 4.3.2, the collection of twice-punctured

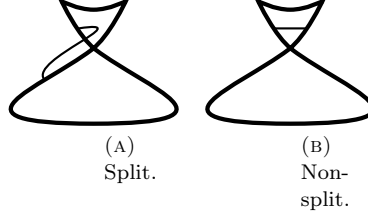


FIGURE 29. Impending merge pairs with linked immersion singles. All cusps are omitted except the pair that come from the initial flip.

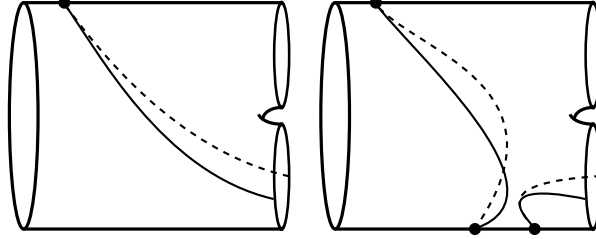


FIGURE 30. Unlinking a genus-decreasing nonsplit immersion single.

critical tori associated to the shift and multislide pairs is embedded, while the intervening critical cylinders are immersed, by  $\alpha$  into  $S^2_{[0,1]}$ . For this reason,  $\alpha_t$  is as required near those values of  $t$  for which  $\text{crit } \alpha_t$  is disconnected and it suffices to restrict attention to the intervening critical cylinders.

**Lemma 8.** One may arrange for  $\iota$  to be a collection of disjoint, embedded circles.

*Proof.* Using Lemma 6, one moves the immersion circles forward and backward in  $t$ , canceling  $R_3$  moves in the critical image by triples of  $R_2$  moves, until the components of  $\iota$  are all disjoint and embedded.  $\square$

**Lemma 9.** Any immersion single that does not bound a disk in  $\text{crit } \alpha$  corresponds to a stabilization.

*Proof.* The immersion single  $\iota$  is a noncontractible simple closed curve in the interior of a critical cylinder  $C =: S^1_{[t_0, t_1]} \subset M_{[0,1]}$  whose immersion locus is exactly  $\iota$ . Viewing  $C$  as a critical base diagram,  $\iota$  is the unique locus of points at which the fiber undergoes surgery; for this reason,  $\iota$  divides  $C$  into a higher-genus and lower-genus side. That is, possibly reversing  $t$ , the fibers along  $C_{t_0}$  are obtained from those along  $C_{t_1}$  by surgery along simple closed curves so that, as  $t$  increases, the base diagrams have the formation, but not the contraction, of swallowtail loops. Preserving the deformation condition, one may cancel any  $R_2$  moves which contribute local minima to  $T|_\iota$  by an isotopy of the deformation, so that the collection of local minima of  $T|_\iota$  coincides with the collection of swallowtail points. The  $R_2$  moves that remain are necessarily local maxima of  $T|_\iota$ , and their number must



FIGURE 31. Moving a swallowtail into a neighboring region of fold points.

agree with the number of swallowtails by the symmetry of Morse functions on circles, with property  $s_2$  implying there is an even positive number of swallowtails. Denoting the swallowtail locus by  $\sigma \subset \iota$ , the local model for swallowtails requires adjacent arcs of  $\iota \setminus \sigma$  to be mapped to each other, so that, following identifications,  $k > 2$  swallowtails would lead to a one-parameter family of points where the deformation is  $k$ -to-one on its critical locus, contradicting the deformation condition. Thus there must always have been exactly two swallowtail points. Now the immersion single corresponds to a deformation that resembles that of a stabilization, except the swallowtails may not lie on the same fold arc. This is easily remedied using the methods of Section 4.1 as depicted in Figure 31, in which a swallowtail jumps over a cusp arc.  $\square$

The remaining immersion locus consists of those circles that bound disks in  $\text{crit } \alpha$ ; the following lemma shows that these circles can be included or converted into deformations corresponding to the moves in Theorem 1, thereby completing the proof: the genus-increasing immersion circles are eliminated and the genus-decreasing circles correspond to sequences of handleslides according to Lemma 1.

**Lemma 10.**

- (1) Any immersion single that bounds a disk or genus-increasing immersion pair can be converted to be part of a stabilization deformation.
- (2) Nested immersion circles can be merged into a single circle.

*Proof.* For each case in the first assertion, there form particular loops in the critical image: for the immersion single, the loop is the one that bounds what used to be the highest-genus region, while for the  $R_2$  it is the pair of loops on either side of the resulting bigon. In each case, one can perform a flip on each loop and cancel the resulting intersections by an  $R_2$  move, then immediately perform the reverse. In the critical base diagram, the immersion locus now has two noncontractible loops which correspond to a stabilization followed by an inverse stabilization, proving (1).

For the second assertion, there is an outermost immersion pair by condition  $s_2$ , and by (1) it is genus-decreasing. Ordering the family of nested circles by their initial appearance with respect to  $t$ , the next pair may be fully nested, in which both components lie within the disks bounded by the outermost circles, or half-nested, with only one component lying within one of the disks. Half-nested pairs are easily seen to have one genus-increasing circle inside the disk, while the other is genus-decreasing. In base diagrams, immediately upon the appearance of the second circle, it is possible to cancel one intersection point coming from the first circle against one intersection coming from the second by an  $R_2$  move which can

be seen to be valid using Proposition 1, then immediately performing the reverse of that  $R_2$  move. This unifies the two circle pairs into a single circle pair.

Both circles of a fully nested pair are either genus-increasing or genus-decreasing. In the first case, one can perform a cancellation between intersection points from distinct circles like before, with the same unification of immersion pairs, using an  $R_2$  move valid by Proposition 2. Thus all nested circles can be assumed fully nested and genus-decreasing. In that case, it is straightforward to verify that a collection of  $k$  genus-decreasing nested pairs corresponds to a  $k$ -fold iteration of the construction from the proof of Lemma 1, leading to a sequence of handleslides.  $\square$

*Proof of Theorem 1.* The preceding lemmas show that  $\alpha$  can be converted to a sequence of the four moves and their inverses. It remains to ensure that the intervening maps give surface diagrams, which is achieved by making sure the higher fiber genus is always at least three. In the second case of Lemma 7 there appeared a way to ensure this around each shift and multislide, and it remains to do satisfy the genus requirement for the intervening values of  $t$ . Suppose the base diagram of the SPWF  $\alpha_t$  has higher fiber genus less than three. Then  $t$  is contained in a smallest interval of maps that violate the genus requirement. More precisely,  $t$  is contained in an interval  $[t_0, t_1]$  such that an inverse stabilization deformation begins just after  $t_0$  and a stabilization deformation finishes just before  $t_1$ , with higher fiber genus less than three for all values of  $t$  between the two moves. By construction,  $\alpha_{[t_0, t_1]}$  is a deformation that induces a sequential modification to the initial surface diagram  $(\Sigma, \Gamma)_{t_0}$  that begins with an inverse stabilization, then proceeds with handleslides, stabilizations and their inverses, all staying within genus less than three, then finishes with a stabilization that results in the genus three surface diagram  $(\Sigma, \Gamma)_{t_1}$ . The stabilizations and inverse stabilizations break  $[t_0, t_1]$  up into subintervals marked by higher fiber genus, and the endpoints of any lowest-marked subinterval correspond to an inverse stabilization at the beginning and a stabilization at the end. Here follow two claims:

- (1) It is possible to move a handleslide to occur before any inverse stabilization that immediately precedes it, possibly at the expense of adding further handleslides before the inverse stabilization.
- (2) The order of the inverse stabilization immediately followed by a stabilization that bound a lowest-genus interval can be switched to be a stabilization followed by an inverse stabilization, increasing its marking by two.

According to the first claim, one may move each handleslide backward out of such a subinterval, then by the second claim it is possible to increase the genus on that subinterval in a way that decreases the number of lowest-marked subintervals; the result follows inductively. Another way to state the first claim, reversing  $t$ , is to say that one may stabilize before doing any handleslide. More precisely, choosing the vanishing cycle  $b$  to slide over  $d$ , one may arrange for the stabilization to occur on a vanishing cycle other than  $a, b, c$  or  $d$  using the trick of Figure 31, and if in the course of the handleslide  $b$  needed to get across the disk on which the stabilization occurred, it is possible to slide  $b$  over the vanishing cycle corresponding to the topmost circle in Figure 2 twice to achieve that effect. As for the second claim, one may first move the swallowtails of one immersion single into the higher-genus side of the other as described in Section 4.1, then apply Lemma 6 to cancel the resulting intersections.  $\square$

## 5. APPENDIX: VALIDITY OF ISOTOPIES

This remark explains exactly when it is possible to push fold arcs around in base diagrams and gives a few common examples. To describe the situation, suppose there is a purely wrinkled fibration  $f_0: M^4 \rightarrow D^2$  and a continuously embedded disk

$$\Delta: \{z \in \mathbb{C} : |z| \leq 1\} \rightarrow D^2$$

such that, for some fold arc  $\phi_0 \subset M_0$ , the restriction

$$f_0|_{\phi}: \phi_0 \rightarrow \Delta(\{|z| = 1, \operatorname{Re} z \leq 0\})$$

is a diffeomorphism (see Figure 32). It would be useful to understand exactly when there is a deformation  $f_t: M_{[0,1]} \rightarrow D_{[0,1]}^2$  such that the image of each point  $\Delta(z) \in f_0(\phi_0)$  in this fold arc follows the path  $\Delta(t\bar{z} + (1-t)z)$  while the rest of the critical image remains fixed. We give a few standard cases in which a proposed isotopy is always valid. Below, the convention is that the starting point for the  $R_2$  move is a map to the disk whose critical locus is a pair of fold arcs whose images intersect transversely at two points.

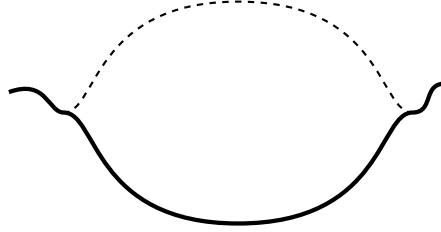


FIGURE 32. A fold arc (in bold) runs along the boundary of a disk  $\Delta$  in the target space of  $f_0$  in preparation to slide across it in some sequence of  $R_2$  or three moves (other critical points omitted).

Suppose, for an  $R_2$  move, that the dotted line  $\Delta(\{|z| = 1, \operatorname{Im} z > 0\})$  in Figure 32 is in the higher-genus side of both fold arcs. Choosing a family of reference points along an arc in the central bigon between the two folds that connects the two intersection points, the reference fibers are decorated with one vanishing cycle and one pair of points coming from the two fold arcs, following vertical paths up and down (respectively) from the reference point, giving a family of Morse functions above the vertical paths. Now it is straightforward to argue that if the pair of points is disjoint from the vanishing cycle in such a Morse function, then the two critical points can be switched by a homotopy. Conversely, if one wants to switch the order of two such critical points in a Morse function, one simply checks for disjointness, which is generic, and switches them, possibly after a small perturbation. For the one-parameter family of such Morse functions obtained by sweeping the vertical paths to the left and right, it is possible that the one-parameter family of vanishing cycles transversely intersects the one-parameter family of pairs of points in the 3-manifold above the one-parameter family of reference points. In the reference fiber, sweeping vertical paths left and right, one sees the vanishing cycle wandering over one or both of the points in a way that may not cancel, much like in the handleslide deformation. In this case, disjointness is not possible for all the vertical paths at once, rendering the  $R_2$  move invalid. If, on the other hand, one may ensure that the

vanishing cycle avoids both points for the durations of its wandering, then the  $R_2$  move is valid, realized as a one-parameter family of homotopies of Morse functions reversing the order of handle attachment. Checking this *disjointness condition* is therefore sufficient to allow the  $R_2$  move.

**Proposition 1.** Such an  $R_2$  move as in the previous paragraph is valid if and only if the disjointness condition is satisfied. In particular, it is valid if the fiber genus inside the central bigon is at least 2 and the vanishing cycles at  $\Delta(\pm 1)$  match (as measured from a point just above  $\Delta(i)$ ).

*Proof.* The first assertion is merely a restatement of the conclusion of the previous paragraph. Denote by  $F$  the generic fiber over a point just above  $\Delta(i)$  and  $F'$  the fiber over a regular value in the central bigon. Also, denote by  $p, p'$  the pair of points in  $F'$  corresponding to the fold arc not containing  $\phi$ , again assumed to be stationary. For the second assertion, because of the matching vanishing cycles, it makes sense to talk about a single vanishing cycle  $v$  of the fold arc containing  $\phi$  that lives in  $F$ , measured near either end of  $\phi$  (and also in  $F'$  for those points actually in  $\phi$ ). Choosing small  $\epsilon > 0$ , the representatives of  $v$  as measured at  $\Delta(\pm 1 \mp \epsilon)$  are isotopic in  $F' \setminus \{p, p'\}$ , which has free fundamental group in which, unlike for the sphere or torus, elements whose representatives differ by crossing  $p, p'$  are distinct. This implies that each crossing of  $v$  over  $p$  or  $p'$  is accompanied by a canceling crossing, which allows the disjointness condition to be satisfied by canceling the crossings.  $\square$

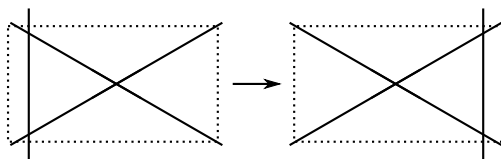
One example of an invalid  $R_2$  move is when the vanishing cycle of the previous paragraph crosses one of the points once. In this case, the fold arc containing  $\phi$  can have different vanishing cycles on either side of the bigon: they differ by sliding over a disk bounded by the other fold arc's vanishing cycle, so that the  $R_2$  move would result in an impossibility:  $\phi$  would be free of intersection points and have vanishing cycles at either end that are not isotopic in  $F$ . Another more subtle example has  $\gamma$  lying within a torus or a sphere. In this case, the one-parameter family of circles traced out by  $\gamma$  can intersect each point of the fiber multiple times with the same orientation; unlike the previous example, the vanishing cycles on either side of the bigon can match, though the  $R_2$  move is still not valid.

**Proposition 2.** Any  $R_2$  move in which the middle region is on the lower-genus side of both arcs is valid.

*Proof.* The paragraph above Proposition 1 still applies, except this time instead of a vanishing cycle wandering around with a pair of points, it is two pairs of points, and 1-manifolds are generically disjoint within 3-manifolds.  $\square$

**Proposition 3.** Any  $R_3$  move such that the central triangle is on the lower-genus side of all three fold arcs is valid.

*Proof.* Consider the neighborhood of the triangle in Figure 33 bounded by the dotted rectangle, foliated by the family of horizontal paths. The restriction of the map to the preimage of each path is a Morse function with three critical points and two or three critical values. The depicted  $R_3$  move can be thought of as affecting the order of critical points of these Morse functions in various ways as can easily be seen from the figure. As remarked above, there is a disjointness condition that must be satisfied on each path if and only if the  $R_3$  move is valid, which in our case is generically satisfied for dimensional reasons as in Proposition 2.  $\square$

FIGURE 33. An  $R_3$  move between three fold arcs.

## REFERENCES

- [AK] S. Akbulut and Ç. Karakurt, Every 4-manifold is BLF, *J. Gökova Geom. Topol.* **2** (2008) 40-82.
- [ADK] D. Auroux, S. Donaldson, and L. Katzarkov, Singular Lefschetz pencils, *Geom. Topol.* **9** (2005), 1043-1114.
- [B1] R. I. Baykur, Topology of broken Lefschetz fibrations and near-symplectic 4-manifolds, *Pac. J. Math* **240** N0. 2 (2009), 201-230.
- [B2] R. I. Baykur, Existence of broken Lefschetz fibrations, *Int. Math. Res. Notices* **2008** (2008).
- [DS] S. K. Donaldson and I. Smith, Lefschetz pencils and the canonical class for symplectic 4-manifolds, *Topology* **42** (2003), 743-785.
- [E] Y. Èliashberg, Surgery of singularities of smooth mappings, *Math. USSR Izv.* **6** (1972), 1302-1326.
- [GK1] D. Gay and R. Kirby, Indefinite Morse 2-functions; broken fibrations and generalizations, (2010 preprint).
- [GK2] D. Gay and R. Kirby, Fiber connected, indefinite Morse 2-functions on connected  $n$ -manifolds, (to appear in *Proc. Nat. Ac. Sci.*)
- [GS] R. Gompf and A. Stipsicz, 4-manifolds and Kirby Calculus, Graduate Studies in Math. **20**, Amer. Math. Soc., Providence, RI 1999.
- [H] K. Hayano, On genus-1 simplified broken Lefschetz fibrations, [arXiv:1012.4049](https://arxiv.org/abs/1012.4049), 2010 preprint.
- [L1] Y. Lekili, Wrinkled fibrations on near-symplectic manifolds, *Geom. Topol.* **13** (2009), 277-318.
- [L2] Y. Lekili, Heegaard Floer homology of broken fibrations over the circle, 2009 preprint.
- [Lev] H. Levine, Elimination of cusps, *Topology* **3**, Suppl. 2 (1965), 263-296.
- [M] J. Morgan, The Seiberg-Witten equations and applications to the topology of smooth four-manifolds, Mathematical Notes **44** (1996), Princeton, NJ: Princeton University Press, pp. viii+128.
- [S] R. Sadykov, Elimination of singularities of smooth mappings of 4-manifolds into 3-manifolds, *Topol. Appl.*, **144** (2004), 173-199.
- [T1] C. Taubes, Counting pseudo-holomorphic submanifolds in dimension 4, *J. Diff. Geom.* **44** (1996), 818-893.
- [T2] C. Taubes, Seiberg-Witten invariants and pseudo-holomorphic subvarieties for self-dual, harmonic 2-forms, *Geom. Topol.* **3** (1999), 167-210.
- [U] M. Usher The Gromov invariant and the Donaldson-Smith standard surface count, *Geom. Topol.* **8** (2004), 565-610.
- [Wa] G. Wassermann, Stability of unfoldings in space and time, *Acta. Math.*, **135** (1975), 58-128.
- [W] J. Williams, The  $h$ -principle for broken Lefschetz fibrations, *Geom. Topol.* **14** (2010), 1015-1061.

DEPARTMENT OF MATHEMATICS, THE UNIVERSITY OF CALIFORNIA AT BERKELEY  
 BERKELEY, CALIFORNIA 94720  
*E-mail address*: jdw@math.berkeley.edu